DOCUMENT RESUME

ED 210 011 IR 009 748

AUTHOR Vaughan, W. S., Jr.; Mavor, Anne S.
TITLE Simulation of a Schema Theory-Based Knowledge Delivery System for Scientists.
INSTITUTION W/V Associates, Annapolis, Md.
SPONS AGENCY National Science Foundation. Washington, D.C. Div. of Information Science and Technology.
PUB DATE May 81
GRANT IST-7904896
NOTE 136p.

EDRS PRICE MF01/PC06 Plus Postage.
DESCRIPTORS *Artificial Intelligence; *Cognitive Processes; *Computer Oriented Programs; Databases; Delivery Systems; Epistemology; *Information Retrieval; Information Seeking; Microbiology; *Models; Online Systems; *Research Tools; Search Strategies
IDENTIFIERS Schema Theory

ABSTRACT

A future, automated, interactive, knowledge delivery system for use by researchers was tested using a manual cognitive model. Conceptualized from schema/frame/script theories in cognitive psychology and artificial intelligence, this hypothetical system was simulated by two psychologists who interacted with four researchers in microbiology to define functional measurements for computer applications. The system worked in real time to provide knowledge delivery services to real research problems through three phases of systems operations--diagnosis, search, and product design. Main systems elements included research paradigms as procedural scripts, information needs as weakly specified frame terminals, and content models as frames. The system's schema enables it to generate representations of information needs as schemata, to plan and conduct targeted searches for relevant information, and to use its inductive and deductive inferencing capabilities. Appendices include interviewing procedures to generate topics and methodological paradigms as well as case studies of actual searches. Nineteen references are included. (Author/RAA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED210011

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

SIMULATION OF A SCHEMA THEORY-BASED KNOWLEDGE DELIVERY SYSTEM FOR SCIENTISTS

W. S. Vaughan, Jr.
Anne S. Mavor

Prepared for:

Division of Information Science and Technology
Directorate for Biological, Behavioral
and Social Sciences
National Science Foundation
Washington, D.C. 10550

Prepared by:

W/V Associates
422 Sixth Street
Annapolis, Maryland 21403

Jerry S. Kidd

Jerry S. Kidd
Principal Investigator
University of Maryland

Technical Report Prepared Under University of Maryland Purchase Order R-264148
National Science Foundation Grant Number IST 7904896

IR009748

May 1981

ACKNOWLEDGMENTS

Dr. Edward C. Weiss, at this time Acting Director, Division of Information Science and Technology, has provided the authors with a strong sense of direction and encouragement to persevere over the several years required for the present conceptualization to evolve.

Dr. Jerry S. Kidd, Professor of Information Science, University of Maryland, has provided the project with long-term technical and administrative support. He has played an important role in identifying potential scientist participants, and in selecting and guiding graduate assistants who carried out many of the information support activities.

Dr. Rita Colwell, Professor of Marine Microbiology, University of Maryland, generously permitted the authors access to her graduate students over several years of exploration into the feasibility of applying schema-like concepts to researcher-information system interaction.

Finally, we most gratefully acknowledge our appreciation of and indebtedness to the microbiologists at Naval Medical Research Institute and University of Maryland who patiently spent many hours with us in our role as "the system".

ABSTRACT

A future, automated, interactive, knowledge-delivery system for use by researchers was conceptualized from schema/frame/script theories in cognitive psychology and artificial intelligence. This hypothetical system was simulated by two psychologists who interacted with four researchers in microbiology as a means to define functional requirements for computer applications. The 'manual' (cognitive) system worked in real time to provide knowledge delivery services to real research problems through three phases of system operations: Diagnosis, Search, and Product Design.

Main elements of the schema/frame/script-based system included the following:

- Research paradigms in a given subfield of science are like procedural scripts which a researcher follows and which a system can learn.
- Information needs are like empty or weakly specified terminals in a researcher's content model or frame.
- A system can use a semi-structured procedure to generate a representation of the researcher's information needs as schemata.
- The system's schema enables it to plan and conduct a targeted search for relevant information, it is not, and need not be, a veridical representation of the researcher's schema.
- A system with inductive and deductive inferencing capabilities can elaborate abstract procedural models from the top down and can construct and elaborate content models in the area of the users information needs.
- Content models are like frames. They are real-world hierarchical structures of known and verifiable facts about content and method in a scientific subfield.
- Content models can be used to hierarchically structure potential knowledge needs in the form of question hierarchies.

TABLE OF CONTENTS

Section		Page
I	INTRODUCTION	1
II	SCHEMA THEORY CONSTRUCTS AND APPLICATION	5
	A. Schemata, Frames and Scripts	5
	B. Simulation Rationalé and Procedure	8
III	SYSTEM REQUIREMENTS FOR MODELS	15
	A. Procedural Models	15
	B. Content Models	21
	C. Summary	26
IV	SYSTEM REQUIREMENTS FOR DIAGNOSIS	27
	A. System Procedure for Diagnosis	27
	1. Step 1: Context for Questions	27
	2. Step 2: Question Sets and Hierarchies	32
	3. Step 3: Context for Answers	34
	B. Summary	38
V	SYSTEM REQUIREMENTS FOR SEARCH	39
	A. System Procedure for Search	39
	B. System Requirements for Indexes to Bibliographic Databases	40
	C. System Requirements for Analysis of Content	42
	1. Recognize Usages and Relationships	43
	2. Recognize Definitions and Synonyms	43
	3. Recognize Members of A Set	44
	4. Recognize Elements of Content Models	45
	5. Recognize Elements of Procedural Scripts	46
	6. Recognize Results, Categories of Study, Etc.	47
	D. System Requirements for Responding to Interdependent Questions	47
	E. Summary	49

TABLE OF CONTENTS (Continued)

Section		Page
VI	SYSTEM REQUIREMENTS FOR PRODUCT DESIGN AND PRESENTATION	51
	A. System Procedure for Product Design	51
	B. Requirements for Researcher-Controlled Output	52
	1. User-Directed Sequence	52
	2. User-Directed Formats	53
	C. Summary	56
VII	REFERENCES	61
	APPENDIX A: PROCEDURE FOR GENERATING A CONTEXT FOR QUESTIONS	A-1
	APPENDIX B: CASE STUDIES	B-1

LIST OF FIGURES

Figure		Page
1	Overview of Procedure for Simulating An Interactive Knowledge Delivery System	13
2	Illustration of A Context for Questions	29
3	Organization of Question Sets Representing Major Areas of Information Needs	33
4	Question Hierarchy Within A Set	35
5	Document Characterization	57
6	Document Synopsis-with-Pointers	58
7	Document Condensation	59

LIST OF TABLES

Table		Page
1	Procedural Model for Taxonomy Research: Main Phases and Tasks	17
2	Procedural Model for Vaccine Development Research: Main Phases and Tasks	19
3	Isolation Methods Model	22
4	Secondary Pathogen Disease Model	22
5	Virulence Factors Model	23
6	An Illustration of How Content Model Elaboration Enabled System to Generate A Question Hierarchy	25
7	Illustrated Procedure for Generating A Context for Questions	30
8	An Illustration of A Context-for-Answers Schema	37
9	Search Results for Pseudomonas Aeruginosa	48

SIMULATION OF A SCHEMA THEORY-BASED KNOWLEDGE DELIVERY SYSTEM FOR SCIENTISTS

I. INTRODUCTION

The present state of automation in science information applications is represented by computer-based systems for retrieving both primary source and bibliographic information. In 1980, approximately 125 bibliographic databases and a few dozen non-bibliographic databases were on-line and commercially available (Hoover, 1980). The bibliographic databases in science and engineering tend to be discipline-oriented, i.e., chemistry, biology, medicine, toxicology; although special-topic systems have been developed in response to critical problems such as energy. Scope of topical coverage and therefore database size vary across systems, but the order of magnitude for coverage is thousands of journals and for size, millions of citations. The stored material consists of bibliographic citations in all cases and may include lists of descriptor words and abstracts. Only a few systems include an on-line document-ordering service and full-text retrieval is currently in an experimental stage.

The bibliographic control systems are currently an important resource for libraries and information centers; they have had a major effect on the training requirements of librarians and on the way libraries collect, retrieve and disseminate information. Their impact on working scientists, however, has been less visible. Research scientists rarely interact with on-line systems as information support devices, and when retrospective searches are required it is typically a librarian who conducts the search and not the scientist (Mavor and Vaughan, 1980). Because the databases are so broad and the librarian-intermediary understands the scientist's information needs at the level of topic-designating words only, searches tend to yield large numbers of citations, many of which are either

irrelevant or redundant. In overview, the current use of automation in science information applications is more to facilitate the archival function of libraries than to efficiently meet the knowledge needs of the research scientist.

One avenue of research in science information systems applications is toward a future, interactive, computer-based system whose commodity is more like knowledge than citations (Weiss, 1977). The general conceptual outline of a future system includes characteristics in sharp contrast to the present state. Databases will be more numerous, each covering material in narrowly-defined content domains. The content of a given database will be information structures representing the current state of development in the domains. The definition of domain may be quite restrictive, e.g., at the level of subfield within a scientific discipline: vaccine development research in microbiology or imagery research in cognitive psychology. System terminals will be located in the research laboratories or offices and the operators will be the scientists who do research in the subfield defined by the system's database. The interaction will be accomplished by a user-directed dialogue whose objective is to transfer to the system the researcher's representation of his current needs for new knowledge. The system will store both procedural and content models typical of research in the subfield and the need-diagnosing process will be accomplished by sophisticated software which implements current schema/frame theories of cognitive psychology and artificial intelligence. Search will be directed toward filling explicit gaps in the researcher's knowledge schema, and the system's output will be new knowledge in the context of the researcher's 'old' knowledge.

The project described in this report is viewed as a primitive first step toward uncovering some of the functional requirements of a schema/frame-theory based system as outlined above. The authors simulated an 'intelligent system' in interaction with working researchers in an unfamiliar field, and

kept track, as best they could, of the requirements they had to satisfy in order to provide effective, knowledge-delivery services. Part II of this report presents the rationalé for a schema/frame-based system and an overview of the approach. Parts III-VI present functional requirements for various system operations in diagnosing knowledge needs, searching for answers, and presenting answer-like material to each of four researchers in microbiology.

II. SCHEMA THEORY CONSTRUCTS AND APPLICATIONS

A. Schemata, Frames and Scripts

Modern cognitive psychology has revived an old theoretical construct, the schema (Bartlett, 1932), as the basic element of memory and the central mediator of cognitive processes; particularly recall and comprehension. (For reviews of modern schema theory see Norman and Bobrow, 1975; Rumelhart, 1975; Rumelhart and Ortony, 1977; Spiro, 1977.) Similarly, and more or less simultaneously, researchers in artificial intelligence (AI) have stressed the need for large information structures, (frames) which integrate both procedural and content knowledge, as context required for computer-based AI systems to 'understand' instructions and relationships in even very limited information domains. (For reviews of frames theory and applications see Minsky, 1975; Kuipers, 1975; Winograd, 1975; Brown and Burton, 1975.) Both schema and frame theories tend to converge in their conceptualizations of how information is represented in memory and what operations might be performed on these representations which explain human comprehension and recall phenomena. Where collaborative research is conducted by cognitive psychologists and AI researchers (e.g., Schank and Abelson, 1977) computer programs are designed and operated as analogues of human cognitive processes. In a sense, computer programs are a means for testing psychological theories of cognitive structure and function.

Scripts are generalized event sequences which a person develops from repeated experiences with instances of the general event or episode (Schank, 1975a, 1975b). Schank's examples are from everyday experience and include both situational scripts as illustrated by a child's birthday party, and procedural scripts as illustrated by the taking of a train to the city. In general scripts are organizing contexts which enable a person

to understand what is happening and to set up expectations about what will happen next once he recognizes that he is 'in' an instance of a particular script. In a sense, a script is a time-extended frame/schema; many of Schank's descriptions of script properties match those of frames/schemata. For example, in all three theories the construct serves as a context of prior knowledge which is used to interpret or comprehend new information; comprehension is defined as a process of searching memory for an existing structure which makes sense of an otherwise novel event. The constructs have both an assimilative and reconfiguring capability; the schema/frame/script is used for comprehending and assimilating new information and is, in turn, modified by the new input. Recall is reconstructive and involves retrieval of partial schema as modified by inferences and intervening events. The constructs impart powerful inferencing and generalizing power; once a portion of a schema/frame/script is 'instantiated', all of the remaining parts can be inferred to be true. The constructs each have 'specialists', or 'demons' or 'framekeepers'; they function to search out empirical evidence that an inferred event or element can be verified. Frames/schemata/scripts have empty 'terminals' or 'slots' which represent gaps or weakly held elements which may be assigned 'default values' by the inferencing mechanisms.

In applying schema theory constructs to the development of a future, interactive knowledge delivery system for use by researchers, we make the following assumptions:

- Doing research is like following a procedural script.

We assume that a large part of a researcher's training and experience can be represented as the accumulation of procedural scripts of the research process as practiced in his subfield. We assume that in all of science, the number of methodological paradigms is far fewer than the number of research topics, and that within a given discipline, or subfield of

a discipline, classes of methods are sufficiently few in number that programs could be constructed which described them all. In psychology, for example, the research topics are large relative to the major classes of methodological paradigm: experimental, correlational, psychometrics, psychophysics, clinical case study, etc. At some level of subfield, the assumption is capable of implementation: if not all of psychology, perhaps all of psychophysics. The point of the assumption, at whatever level it can be applied, is that once a researcher has chosen to approach a topic with a given methodological paradigm, implementation of the method is highly prescribed. There is a series of hierarchical, time-ordered steps to be taken which involve decisions the researcher must resolve in order to progress. Assuming that a computer program 'knows' the researcher's chosen script, the system can locate his progress, identify the decisions which are required now, know that certain prior tasks/steps have already been accomplished, and anticipate next steps.

- **An information need is like an empty or weakly specified 'terminal' in the researcher's schema/frame.**

As the research progresses through its various phases, the researcher is required to make judgments and resolve decisions in order to accomplish the required tasks/steps. At these junctures, the researcher brings his knowledge to bear on the problem. Where he experiences knowledge gaps or uncertainties he identifies new information requirements. These gaps occur in the context of the researcher's already-structured schema or complex of interrelated schemata. To the extent that the researcher's current knowledge structures can be approximated by the system, the new knowledge requirement can be precisely specified. We assume that a diagnostic procedure can be developed by which a computer program interrogates a researcher to create information structures which approximate

the researcher's schemata, identifies information gaps, and specifies the characteristics of information that satisfies gaps.

B. Simulation Rationalé and Procedure

In the present project, a two-person team of scientists trained in experimental psychology simulated a future interactive system as a means to identify functional requirements for its design. The strategy was for this team to simulate a 'smart' software system in conversational interaction with researchers who had current information needs, and in doing so to record the functional requirements the system had to satisfy in order to provide effective information support. This 'system' interacted with four researchers in microbiology who were willing to participate in the interactive sessions for the value of the information the system promised to deliver. Three of the participants were senior researchers at a Navy laboratory involved in vaccine development, each with a different organism and at a different stage in the process. The fourth researcher was a university Ph. D. candidate doing thesis work on the taxonomy of selected marine microorganisms.

The 'system' personnel were in no sense a 'tabula rasa'. Both had strong backgrounds in research methods in psychology generally and experimental methods in particular. Our ability to comprehend and model research procedures in vaccine development and taxonomy was due in large part to the strong methodological analogies of the former to factorial experimental design, and of the latter to discriminant function analysis. Both team members were less than naïve with respect to the content; we certainly knew more than the average citizen since the previous year had been spent pre-testing parts of the simulation idea with graduate students working on various problems in marine microbiology. We had not encountered either the taxonomy or the vaccine development model, but we knew an unspecifiably-large amount of content relative to a naïve system. Our attempts

to specify the procedural and content requirements of a future interactive system were almost certainly underestimates in both areas; on the other hand, a 'system' with no schema/frame, probably could not have interacted at all.

The procedure of simulating a computer system by using human substitutes might seem peculiar to some readers because contemporary practice in artificial intelligence research is to use the computer system to simulate human behavior. However, the present procedural orientation has a relatively long and fruitful history. In the early days of computer system development, it was often possible to conceive of a function that a computer could potentially fulfill but not have either the CPU or memory capacity much less the program to actually perform the function. The basic question for research was: Is it worthwhile to build the capacity and create the programming? Parsons (1972) describes a series of experiments done in 1962 that followed this approach and successfully distinguished between several alternative functional specifications for a computer-aided command and control system. Similar methods were employed in a series of studies of decision making in complex organizational settings (Kennedy, 1962). The orientation in the present instance was much the same. The functions conceived for an interactive, computer-based information system have not yet been translated into working computer programs. The intent was to make a preliminary determination of whether the effort to formulate such programs could be a worthwhile enterprise.

Our initial plan was to concentrate exclusively on defining a future system's functional requirements for information need diagnosis: the front-end of a knowledge delivery system. Requirements for need diagnosis resulting from the project would not be tied to the current techniques and mechanisms for information storage and retrieval as would results of follow-on activities. In practice, however, real scientists actively engaged

in important research are not inclined to stop at having their information needs diagnosed, and so the 'system' simulated remaining phases of searching for and retrieving documents, extracting information from documents and preparing information products for evaluation by each researcher. A typical simulation involved several interactive work sessions interrupted by time delays during which the 'system' searched and screened citation lists, tracked-down journal articles from a library, etc.; the time frame covering several weeks.

The general procedure was as follows: the 'system' and researcher held a first meeting to define researcher information needs using a semi-structured procedure. This interaction varied considerably on the dimension of dialogue control. The system had a prescribed set of information categories to learn about as a context for understanding the researcher's needs, and in some cases the interaction was driven by the system proceeding through its list of topics, in others the researcher took control and the system filled-in appropriate information categories. The system reviewed what it had learned from the diagnostic interaction and organized the researcher's information needs into sets of questions. A second interactive session was held wherein the researcher was asked to review the system's conception of his information needs as question sets and to provide an approximation of his knowledge structure or schema for each question. Next, the system planned a key word-oriented search strategy with appropriate bibliographic systems, typically MEDLINE and BIOSIS, to generate citation lists. The system screened these citations for relevance to the researcher's questions and selected a small subset (less than 20) of the most clearly relevant. The question set/citation list combinations were reviewed by the researcher in a third interactive session during which the researcher confirmed the relevance of the system's selections and eliminated those with which he was already familiar. Then the system located the documents

for the few remaining citations (usually 6-8), read them, extracted information that appeared relevant to the researcher's questions and prepared informational materials in a variety of formats for presentation to the researcher. These products were reviewed by the researcher in a fourth interactive session, mainly to evaluate the informational content of the materials and to suggest changes in their format or level of descriptive generality/specificity which would better suit the researcher's purposes. The document-reading and product-design steps were recycled for each of the researchers' questions so that product evaluations for Question #1 led to modifications in product design for Question #2. The interactive procedure is illustrated in Figure 1.

From the system's point of view, the simulation of the automated knowledge delivery system progressed through three main phases: Diagnosis, Search, and Product Design. How we did each phase and what we learned about future system requirements from examining our experiences in doing them are described in the following sections, III-VI. Appendix B contains the four specific case study reports on which the generalizations of these sections are based.

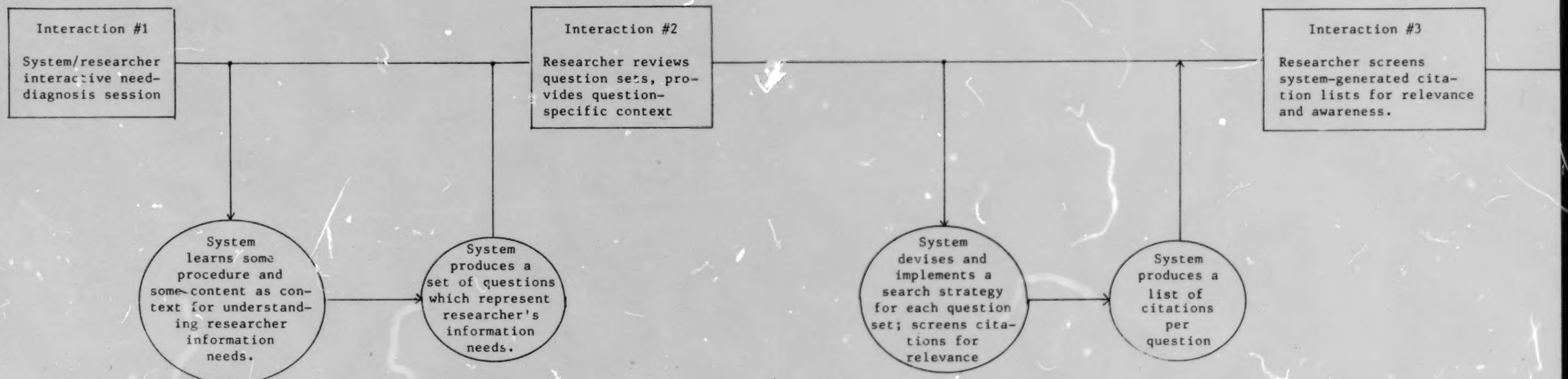
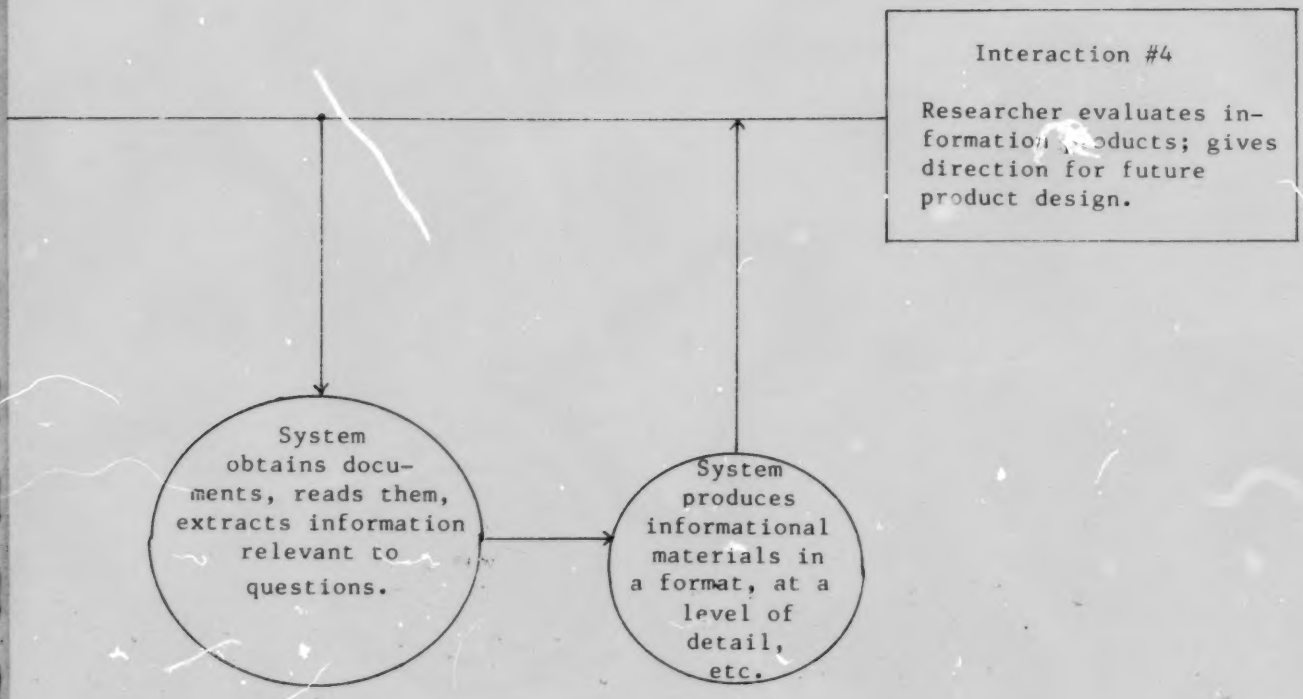

Interaction #1
System/researcher interactive need-diagnosis session
Interaction #2
Researcher reviews question sets, provides question-specific context
Interaction #3
Researcher screens system-generated citation lists for relevance and awareness.
System learns some procedure and some content as context for understanding researcher information needs.
System produces a set of questions which represent researcher's information needs.
System devises and implements a search strategy for each question set; screens citations for relevance
System produces a list of citations per question

Figure 1. Overview of Procedure for Simulating An Interactive Knowledge Delivery System


Interaction #4
Researcher evaluates information products; gives direction for future product design.
System obtains documents, reads them, extracts information relevant to questions.
System produces informational materials in a format, at a level of detail, etc.

III. SYSTEM REQUIREMENTS FOR MODELS

Implementation of a schema-theory based approach to an automated knowledge delivery system required both procedural and content models as context for system-researcher interactions during Diagnosis and for system-database interactions during Search. In the present application, the system initiated diagnostic interactions using a procedural model or 'script' appropriate to the researcher's approach as a framework for attaching content information about the researcher's topic. As the system-researcher interaction proceeded, the system expanded the procedural model hierarchically 'downward' to the level of specificity of the task or step that defined the researcher's current state of progress and developed one or more content models to the extent necessary as a context for the system to comprehend the researcher's discussion of his knowledge needs. During Search, the system used the elaborated content models to recognize relevant information for extraction and delivery to the researcher.

A. Procedural Models

Two general procedural models were constructed to represent two research paradigms in microbiology: taxonomy research (Table 1)and vaccine development research (Table 2). Each model was developed hierarchically from the most general level, Phases, to a second level, Tasks within phases. The models were intended to describe the operations required to do research in taxonomy and in vaccine development; they defined how research is done without reference to specific organisms of concern to the research.

The system successfully used these rather weak procedural models as a point of departure for Diagnosis. Each researcher was able to locate his current research activity within the Phase/Task structure of the model, and begin the discussion of his content problems in context of the task to be performed; i.e., the procedural model served as a frame for initial content model construction. The system was able to expand its initial procedural models in the area where the researcher was working. As can be seen from the vaccine development research model of Table 2, the system learned least about the task structure in the Antigen Test and Evaluation phase; none of our researcher participants was working at that stage. Where the information needs of the researcher included content about methods, the system was able to expand the Phase/Task model in taxonomy research to the level of steps within tasks as shown by taking a section out of Table 1, as follows:

Procedural Model for Taxonomy Research

Phase 2. Strain Selection

Task E. Outline Procedures for Isolating Desired Strains

Step 1. Select An Isolation Method

Step 2. Select A Recovery Medium

Step 3. Select or Modify Isolation Apparatus for Field Applications

Table 1. Procedural Model for Taxonomy Research: Main Phases and Tasks

Phase 1. Problem Selection
Tasks
A. Identify an organism as not classified at the species level.
B. Determine the importance of studying the organism.
C. Determine the feasibility of studying the organism.
Phase 2. Strain Selection
Tasks
A. Determine the group or groups from which strains are to be sampled.
B. Determine where strains should be sampled from and how many strains should be used.
C. Develop a sampling plan.
D. Outline procedures for collecting samples.
E. Outline procedures for isolating strains.
F. Collect sample.
G. Isolate strains.
H. Receive and analyze samples from other investigators.
Phase 3. Test Selection
Tasks
A. Determine types of tests to be performed: serological, biochemical, morphological.
B. Identify tests performed by other investigators or standard tests that might be useful.
C. Select tests to be performed from existing tests based on information on: • Discrimination • Precision • Reliability • Feasibility
D. Design/develop new tests to be performed.
E. Perform tests to see if they meet criteria of discriminability, etc.
F. Reduce number of tests to those that perform required discriminations for classification.

Table 1. Procedural Model for Taxonomy Research: Main Phases and Tasks (Continued)

Phase 4. Test Implementation
Tasks
A. Outline procedures/methods/equipment requirements of each test.
B. Grow organisms and separate into groups for test applications.
C. Perform tests.
D. Record results.
Phase 5. Cluster Analysis of Test Results
Tasks
A. Organize results into a format for computer processing using Numerical Taxonomy Program.
B. Input data into computer and obtain printouts of cluster analysis.
C. Input test data on other organisms in the genus which were collected and provided by outside investigators.
Phase 6. Interpretation of Clusters
Tasks
A. Analyze clusters to determine relatedness between test organism and other organisms in the data base.
B. Determine if relatedness is high enough to consider test organism representative of a cluster.
C. If test organism is related to a cluster containing a known species, name the organism.
D. If organism is not related to a known species check other sources, use other classification systems, publish description in professional literature.

Table 2. Procedural Model for Vaccine Development Research: Main Phases and Tasks

Phase 1. Problem Selection
Tasks
A. Select organism to study and identify species or strains which appear to be disease correlates.
B. Establish importance of studying the selected organism based on types of disease caused, disease severity, and existing treatment problems.
Phase 2. Culture Requirements Determination
Tasks
A. Obtain organisms from other laboratories/investigators.
B. Isolate the organism; develop/learn procedures for identifying and isolating organism either from host cells or from the environment.
C. Grow organisms for experimentation; develop/learn organism's growth media requirements, i.e., the compounds metabolized by the organism.
Phase 3. Organism Characterization
Tasks
A. Describe the organism's morphological characteristics through electronmicroscopy; determine sizes, chapes; presence of attachment structures, identify and characerize attachment structures.
B. Describe organism's biochemical characteristics; develop/learn appropriate tests.
C. Describe toxic products produced by the organism; develop learn assays and tests for determining presence and amount of each selected toxic product and specify the chemical structure of each toxin identified.
D. Identify strains which produce various combinations of toxic products.
Phase 4. Virulence Factors Determination
Tasks
A. Examine each toxic product in terms of penetration, attachment, destruction of host cells; test strains with varying combinations of toxic products on host cells.
B. Examine the role of attachment structures (pili) in establishing the bacteria in a host cell; compare organisms with and without pili and measure the level of disease caused in the host cell.

Table 2. Procedural Model for Vaccine Development Research: Main Phases and Tasks (Continued)

Phase 4. Virulence Factors Determination (continued)
Tasks
C. Determine the biochemical characteristics of pili (or any effective attachment structure) and identify substances that destroy these structures; develop/learn/conduct biochemical analysis.
D. Select/develop animal models for testing impact of various toxins (singly and in combination).
E. Determine disease levels and organ damage caused by various combinations of toxic products; develop experiments testing different dosages under different conditions with one or more animal models.
F. Describe organ damage due to presence of organism in human hosts.
Phase 5. Antigen Test and Evaluation
Tasks
A. Develop antigen(s) which will act against the toxins produced by the organism.
B. Conduct experiments to determine safe but effective dosages of antigen.

B. Content Models

Initial system-researcher interaction in need diagnosis involved the piecemeal construction of content structures which defined the topics of concern to the researcher within a given research paradigm. The vaccine development research model, for example, was used with three researchers each studying a different organism in a different phase. One researcher was exploring growth media requirements for Rickettsia, a second researcher was determining the enzyme product characteristics of Aeromonas, Campylobacter and Yersinia, and a third was doing experiments to determine virulence factors of Pseudomonas aeruginosa. The procedural model enabled the system to comprehend what the researcher was trying to do (tasks), and content models were required for the system to understand the organisms and phenomena which were the topics of concern in doing the tasks. The content models were required as context for comprehending what the researcher knew and needed to know for the purposes of his current phase/task/step. The procedural models enabled the system to comprehend reasons why the researcher needed knowledge, i.e., its task application; the content models were required to specify the area and level of detail of his knowledge needs.

Three examples of content models are shown as Tables 3, 4, and 5. Table 3. Isolation Methods Model, was developed as a context for identifying the knowledge needs of the researcher who planned to isolate aquatic strains of streptococcus from Chesapeake Bay. Table 4. Secondary Pathogens Disease Model, provided a general context for understanding the requirements of two researchers who were working with secondary or opportunistic pathogens (pseudomonas and aeromonas). Table 5. Virulence Factors Model, enabled the system to comprehend needs for information about how the enzyme products of bacteria contribute to disease development.

Table 3. Isolation Methods Model

- There are several methods for isolating a bacterial genus (e.g., streptococcus) from a sample of water; these methods are usually called 'tests' and include Membrane Filter Test, Most Probable Number Test, etc.
- All isolation methods include the use of a recovery medium; a recovery medium is a combination of compounds designed to select for the organism; an effective medium recovers only the target organisms and rejects all others.
- Compounds composing the recovery media are in the form of broths or agars and are often labeled by letters which designate the ingredients of the compound (e.g., PSE media, PSE broth, PSE agar are synonymous).
- A range of media can be used with a given isolation method; the selection depends on the organism to be captured.
- In addition to selectivity in recovering the target organism, other criteria for assessing isolation methods include availability of apparatus and materials, simplicity of application, and adaptiveness to in-the-field sampling conditions.

Table 4. Secondary Pathogen Disease Model

- Secondary (opportunistic) pathogens are bacteria harmless to healthy persons, but which attack seriously ill (cancer, tuberculosis, pneumonia) or traumatized patients (deep wounds, amputees, burn victims).
- They are common to most environments, but are particularly prevalent in hospitals.
- They establish a colony in the patient's body (wound, burn, respiratory system, renal system).
- They eventually penetrate the patient's bloodstream. This condition is called bacteremia (septicemia, sepsis, blood poisoning).
- The bacteremic patient may die of septic shock and failure of a critical organ (heart, lungs, kidneys, liver).
- The incidence of bacteremia and the mortality rate associated with bacteremia may vary systematically with the patient's initial clinical illness/trauma.

Table 5. Virulence Factors Model

- Microorganisms have structural characteristics and produce enzymes which may have a role in disease development.
- Pathogenic roles include penetration of the host cell, attachment and damage; each role may be accomplished by one or a combination of enzymes or other characteristics of the microorganism.
- Some microorganisms have pili (tentacle-like structures) which may enable the organism to attach itself to host cells.
- Some enzyme products of microorganisms include protease, elastase, hemolysin, endotoxin, exotoxin.
- Studies with pseudomonas aeruginosa suggest that protease has a role in penetration; exotoxin A, a role in cell damage.
- How a microorganism accomplishes a given role in disease development is called mechanism of action; e.g., exotoxin A from pseudomonas aeruginosa appears to damage host cells by preventing the cell from synthesizing protein.
- Mechanism of action may be specific to the type of host cell under attack by the microorganism; e.g., pseudomonas aeruginosa appears to penetrate lung tissue by neutralizing the aveolar microphage, organisms in lung cells which attack bacteria.
- Within a bacterial specie there are several subspecies or strains. The strains differ in their enzyme products and in their structural characteristics (e.g., presence or absence of pili). Enzyme products vary in characteristics such as amount, concentration, and pyocine type; these variations may be related to virulence differences among bacterial strains.

The three content models are shown in Figures 3, 4 and 5 as lists of facts about each content domain. The content models are hierarchical, however, and their structure enabled the system to generate parallel question hierarchies and thereby to more precisely specify the researcher's needs. For example, two researchers in vaccine development had information needs in the area of virulence factors; one at a very early state in the procedural script for virulence factors determination, the other at a more advanced stage in the script. Both expressed their information requirements at the general level, i.e., "I need information about virulence factors in P. aeruginosa (A. hydrophila)", but the kinds of answers each required were at different levels in the content hierarchy. In order to comprehend the differences in their requirements, the system constructed a content model for Virulence Factors which revealed its hierarchical structure. Table 6 is an illustration of a partial content hierarchy for 'Virulence Factors' and a parallel hierarchy of questions which enabled the system to more precisely specify the nature of the answers sought by each researcher.

Table 6. An Illustration of How Content Model Elaboration Enabled System to Generate A Question Hierarchy

	Content Model Hierarchy for Virulence Factors	Researcher Question Hierarchy
Level A: Descriptive	Pathogens have structural features which can be observed, and they produce enzymes that can be assayed (measured). • Structures: pili • Enzymes: exotoxin endotoxin protease elastase hemolysin	What features/enzyme products of the organism have been observed/ assayed which might contribute to disease development?
Level B: Relational; Qualitative	There are three pathogenic roles in disease development: cell penetration, attachment and destruction. There is research evidence that certain features/enzyme products accomplish specific roles. • Pili accomplish attachment • Exotoxin accomplishes cell destruction • Protease accomplishes cell penetration	Which features/enzyme products of the organism contribute to which roles in disease development ?
Level C: Mechanism of Action: Qualitative	Exotoxin A from P. aeruginosa acts like diphtheria toxin. It destroys the host cell by inhibiting protein synthesis. Protease from P. aeruginosa appears to penetrate lung cells by neutralizing aveolar microphage. Aveolar microphage are organisms in lung cells which attack bacteria.	By what means does the pathogen's features/ enzyme products accomplish a given pathogenic role?
Level D: Mechanism of Action: Quantitative	Enzyme products can be assayed for several characteristics. • Concentration (titre) • Pyocine type Hemolysin in high concentrations is destructive of corneal tissue; hemolysin in low concentrations is not.	How much of or what characteristic of the enzyme product accomplishes the pathogenic role?

C. Summary

Procedural models are an effective and efficient means to initiate system-researcher interaction in need diagnosis. They effectively locate the researcher in a paradigm appropriate to his field; they focus diagnostic interaction on those content areas which embed the researcher's knowledge needs.

Procedural models are generalizable over a range of topics. Vaccine development research proceeds through a prescribed series of steps regardless of the target organism.

Both procedural and content models are hierarchical and capable of both lateral and vertical expansion.

A knowledge delivery system requires only partial models and mechanisms for expanding models in the areas of researcher-specified needs. The system can store fragments of content models, such as virulence factors, and interactively develop the partial model according to the specific area and level of detail which embeds an individual researcher's knowledge needs.

A hierarchical content model can be used to generate a parallel question hierarchy, enabling a system to precisely define the level of answers a researcher seeks.

IV. SYSTEM REQUIREMENTS FOR DIAGNOSIS

The system's objective in Diagnosis was to specify the researcher's needs for new knowledge as well-defined gaps in a schema representation. The process involved the successive development of content models to a level of detail which contained the researcher's needs. Since the system initiated Diagnosis with only general content knowledge, the procedure required three steps; these are described and illustrated in the following sections.

A. System Procedure in Diagnosis

1. Step 1: Context for Questions

In Step 1, the system used the phase/task procedural model as a vehicle for structuring a system-researcher discourse on the content of the researcher's problem. The discourse was semi-structured in that the system possessed a series of questions (see Appendix A) to ask the researcher which were intended to develop the content, but the course of the interaction was never that constrained. The system asked the first question and the researcher responded to it. What happened next depended on how far beyond the first answer the researcher was inclined to continue the discourse. In one of the four interactions, the system tried to implement a rigidly-systematic, question one-answer one protocol; this was possible but not comfortable. It overly constrained the researcher and interrupted his own ideas regarding the logical development of his problem. However, since the system knew the answers it needed and could recognize answers to yet unasked questions, the semi-structured procedure was in all cases successful in eliciting content information sufficient for the system to comprehend the researcher's information needs at a gross level. This gross level of comprehension could be variously labeled (Topics, Issues, Areas, Problems,

etc.). We arbitarily decided to call this level of representation Questions and the purpose of Step 1 as generating a context which permitted their recognition.

The general outline of the procedure for generating a context for questions is diagrammed in Figure 2 and illustrated in Table 7 for the case of the researcher working on taxonomy of marine streptococcus. The categories of information used to provide a context for questions are:

- Statement of the Research Topic: name of the organism of interest; phenomenon of the organism being researched; importance of the research, justification or rationalé.

- Researcher's Progress in the Procedural Script Applicable to His Work: current phase/task; goal hierarchy.

- Characterization of the Researcher's Experience in the Topic Area: level of sophistication.

- Characterization of the Status of Research in the Topic Area (General Background): general status of the problem; approaches to the problem; current obstacles, impediments to progress.

- Characterization of the Researcher's Current Problem: problem statement; rationalé for research approach; level of implementation of the approach; current areas of information needs.

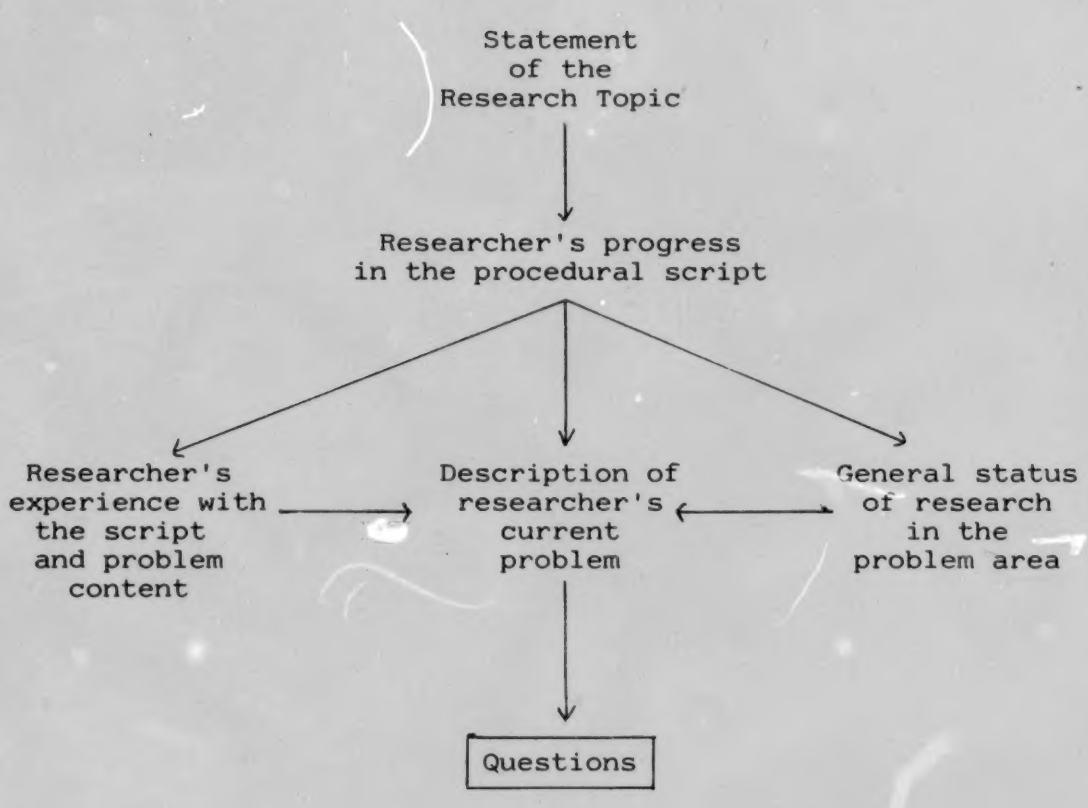

Statement of the Research Topic
Researcher's progress in the procedural script
Researcher's experience with the script and problem content
Description of researcher's current problem
General status of research in the problem area
Questions

Figure 2. Illustration of A Context for Questions

Table 7. Illustrated Procedure for Generating A Context for Questions

Question	Answer
Step 1. Statement of the Research Topic	
1. What is the name of the organism(s) of interest?	• Streptococci and staphylococci.
2. What is being studied about the organism(s)?	• Isolation and taxonomy of clinical and aquatic strains.
3. What is the importance of this research?	• Streptococci and staphylococci in aquatic environment may pose a threat to the health of humans and commercially valuable seafood. • Streptococci and staphylococci may be useful indicators of fecal pollution in aquatic environments.
Step 2. Location of Researcher in Appropriate Procedural Script	
1. Which research paradigm are you using to study this problem?	• Taxonomy in clinical microbiology.
2. Do the phases in your research match the phases shown in this taxonomy script?	• Yes
3. What phase of your project are you currently working on?	• Phase 6. Cluster Interpretation of Clinical Strains • Phase 2. Strain Selection for Aquatic Strains
4. Which of the tasks under strain selection are you currently working on?	• Procedures for collecting samples from aquatic environment (Phase II, Task D). • Procedures for isolating strains from aquatic environment (Phase II, Task E).
Step 3. Characterization of the Researcher's Experience with the Script and Problem Content	
1. Have you worked in taxonomy research before?	• Yes. All of my work has been with clinical strains. • Work has been done on classifying a variety of clinical species. • The phases and tasks of the taxonomy paradigm have been performed many times.

Table 7. Illustrated Procedure for Generating A Context for Questions (Continued)

Question	Answer
Step 4. Characterization of the General Status of Research in the Area	
1. What do you know about the research that has been conducted in this area?	• Classifications of clinical streptococci. • Tests for classifying streptococci. • Tests for isolating clinical samples. • One test, Membrane Filter, for isolating streptococci from aquatic environments; this test has drawbacks.
Step 5. Characterization of the Researcher's Current Problem	
1. What is your research approach?	• Use experience with clinical strains as an analog for working with aquatic strains? • Use tests that maximally discriminate clinical strains on aquatic strains.
2. What are your current problems?	• Determining an effective, efficient way to isolate streptococci and staphylococci from an aquatic environment.

2. Step 2. Question Sets and Hierarchies

In Step 2, the system analyzed the content information and lists of questions it had acquired in Step 1 for the purpose of generating logically-related subsets of questions; i.e., Question Sets. The criterion for this procedure was to create sets which were maximally independent of one another, enabling the system to develop independent searches of the literature (database). We view this process as analogous to the idea of decomposing the problem space into manageable components (Simon, 1969). The analogy to problem-solving strategy required a minor modification, for although the researcher has the research problem and the information needs, it is the system which does the decomposing and imposes the criteria for 'manageability'. The system eventually will search for answers from its database, and so it must generate a representation of the researcher's information needs which facilitates/enables the development of search strategies. We conceptualize the outcome of Step 2, Question Sets, as a representation of the researcher's information-need domain which has functional value for directing system search activities, and not as a veridical representation of the researcher's 'mental model' of his research problem. Our representations cannot reproduce the researcher's 'true' internal schemata in whose terms he thinks about his work; these are assumed to be far too complex to model and probably impossible to elicit. Furthermore, veridical representation is unnecessary, the system does not need to 'know' all that the researcher knows about the problem, and the mapping from the researcher to the system does not need to satisfy strong requirements; it is only necessary for the system's representation to be adequate for a targeted information search.

In the present simulations of a future system, the Question Sets generated during Step 2 were coincident with the content models which were outlined in Step 1. The researcher's information needs were understood by the system as questions embedded in one or more content models which the system perceived to be independent. Figure 3 illustrates the outcome of Step 2

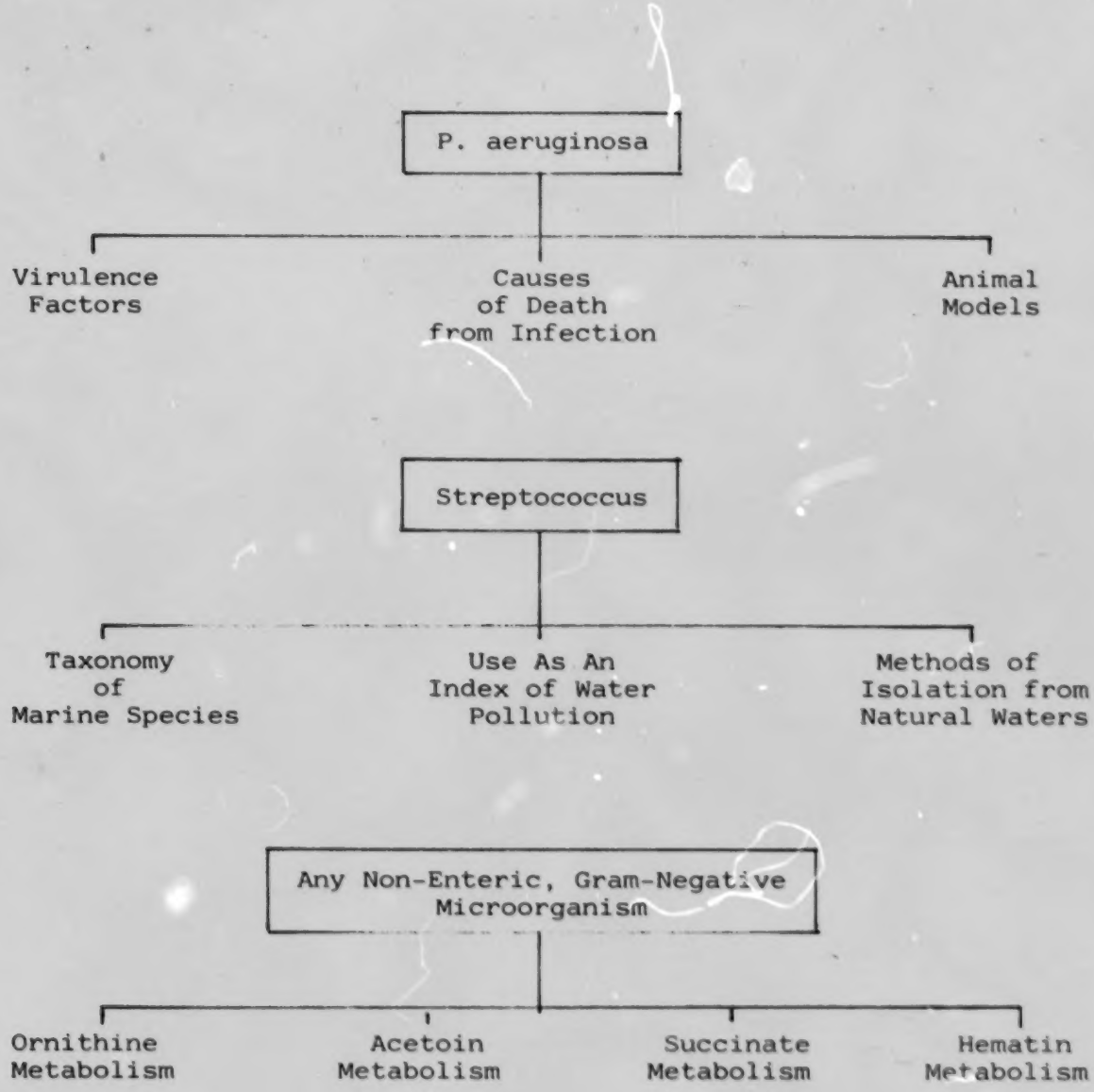


Figure 3. Organization of Question Sets Representing Major Areas of Information Needs

activities for three of the reserachers. In each case the Question Sets were more or less independent aspects of an organism or class of organisms which encompassed the researcher's information needs.

The system's initial structuring of the Question Sets was at the level of lists, which tended to reflect our weak level of development of the content models from which the questions were derived. As we developed each content model more deeply, we were able to comparably develop the structure of the questions within each Question Set; and since the content models were hierarchical, the questions within a set were hierarchical. The parallel development of content model and question hierarchies was illustrated in Table 6 of Section III for 'Virulence Factors'. The process of hierarchical expansion occurred in stages as the system 'learned' from interacting with the researcher during Diagnosis and with the database during Search. Figure 4 is an illustration of an hierarchy of questions within another of the content models: Methods of Isolation from Natural Waters. The value of the hierarchical expansion of Question Sets, of course, was to increase the precision of the search for answers.

3. Step 3. Context for Answers

The purpose of Step 3 was to produce a knowledge structure, i.e., schema representation, for each of the questions defining a researcher's information needs. The schema representations were to define knowledge the researcher already possessed, and to identify gaps he knew needed to be filled. Gap definition was to insure relevance, and current knowledge definition was to avoid redundancy of materials retrieved in searching for answers.

In practice, the system/researcher interaction was successful in hierarchically elaborating both the question sets and

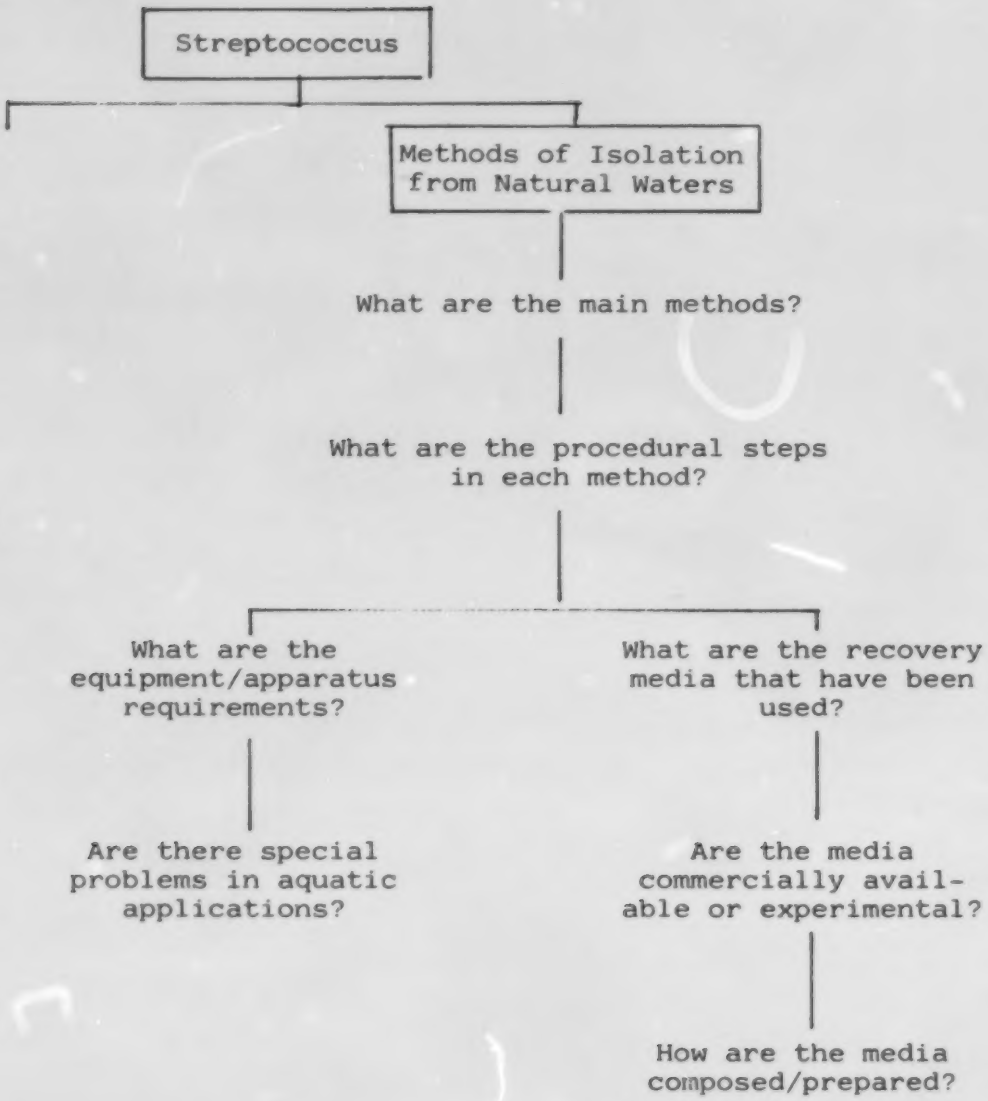

Streptococcus
Methods of Isolation from Natural Waters
What are the main methods?
What are the procedural steps in each method?
What are the equipment/apparatus requirements?
What are the recovery media that have been used?
Are there special problems in aquatic applications?
Are the media commercially available or experimental?
How are the media composed/prepared?

Figure 4. Question Hierarchy Within A Set

the content models to the level of the researcher's knowledge needs. The interaction was much less successful at specifying prior knowledge as a basis for excluding redundant material. The system was consistently stronger at defining relevance that it was at defining redundancy. Either the researchers found it easier to talk about their requirements than about what they already knew, or the system was weak at recognizing when it was important to specify prior knowledge. The former may be so; the latter certainly was. We consistently had the experience of reviewing documents that contained information we knew for sure was relevant, but were uncertain about redundancy. In instance after instance we allowed the researcher to tell us he was interested in learning about 'new' X's or 'improved' Y's and failed to specify 'old' and 'unimproved'. For example, one researcher told us a great deal about his interest in learning about methods for isolating streptococcus from natural water samples; Table 8 illustrates our attempt at schema representation. He specified one method: Membrane Filter Test. Our literature review turned up five isolation methods used in combination with sixteen different recovery media; all of relevance to the question set, but which of it was 'new' to the researcher? Our solution was to present 'answers' hierarchically; i.e., first a 'menu' of method names, then details of those selected by the researcher as 'new'.

A future system will benefit from strong routines which recognize requirements for current knowledge definition, and then ask the researcher to specify 'old' before the system searches for 'new'.

Table 8. An Illustration of A Context-for-Answers Schema

Phase 2: Strain Selection Task E: Outline Procedures for Isolating Strains		Question Set: What isolation methods have been used in aquatic environments?
Facts	**Inferences**	**Gaps**
One method for isolating fecal streptocci is the Membrane Filter Test The Membrane Filter Test appears unreliable in natural waters. Isolation methods involve a procedure for selecting one organism and not others Media are used to selectively recover one organism and not others Media are composed of combinations of compounds	Material suspended in natural water may pose problems for traditionally clinical isolation methods. Other methods may be better for this application.	• What studies show the value of the Membrane Filter Test • What other methods have been used to isolate gram-positive cocci from an aquatic environment; from sediment and debris. • How do these methods compare with the Membrane Filter Test

B. Summary

The system was able to construct schema/frame-like content models using a semi-structured, interactive procedure. The procedure was capable of implementation in both system-controlled and researcher-controlled dialogue styles, but the researcher-controlled style was strongly preferred.

Content models encompassed a hierarchical set of questions in the area of the researcher's knowledge needs; the development of either hierarchy enabled the elaboration of the other.

The system arbitrarily decomposed the researcher's information need domain into Question Sets according to a pragmatic criterion enabling it to plan minimally overlapping searches of its bibliographic database. Maximally independent Question Sets were structured according to the system's perception of the 'independent' content models which contained them.

A question schema enabled the system to specify relevant answers with great precision, but failed to specify redundant answers; the system requires stronger procedures for specifying the researcher's current knowledge at a detailed level.

In the absence of strong schema detail about facts known to the researcher, the system can present its knowledge hierarchically and sequentially according to researcher selections of routes he chooses to pursue.

V. SYSTEM REQUIREMENTS FOR SEARCH

A. System Procedure for Search

The system's ultimate objective was to provide the researcher with information products in appropriate formats which satisfied his knowledge needs; i.e., filled gaps in his schema. The purpose of the search phase was to provide the raw materials for these products by identifying information from the scientific literature which matched the schema representation.

Starting with the researcher's question schema the system performed the following steps:

- Select combinations of key terms which best represent the question, query the database with these terms and obtain citation lists.
- Review citation lists for relevance to the question.
- Have researcher review selected citations for both relevance and prior awareness.
- Retrieve citations selected by the user, perform content analysis and select relevant information.

This procedural sequence was implemented for thirteen question schemata. The value of the schema approach to need diagnosis was evidenced by the results of the researchers' review of the citation lists compiled by the system. Of 170 citations shown to the four researchers as a check on relevance, only two were judged irrelevant by the researchers; a retrieval precision of 99%.

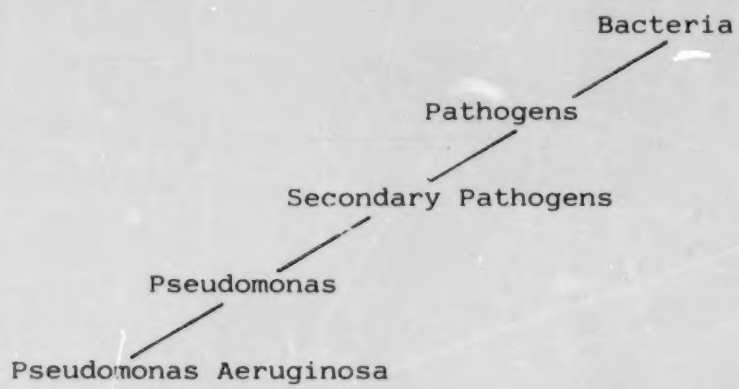
An analysis of the processes performed and the problems encountered in each search step led to the identification of general requirements for a future system. These requirements are in three areas:

- Indexes to databases.
- Content analysis of titles and lines of text.
- Response to content-interdependent questions.

The following sections present our findings in each of these areas.

B. System Requirements for Indexes to Bibliographic Databases

The controlled vocabulary of automated bibliographic databases provides a hierarchical subject classification in the form of key words and concept codes. This classification includes, at the lowest level, the name of a specific organism, process, product or method and at succeeding levels, sets of progressively comprehensive categories. For example, index terms could be provided for:



Citations classified at one level in the hierarchy may or may not be classified at other levels; classification tends to be at the most specific level. Selection of terms and term combinations results in lists of citations which include these terms either in the title or in the abstract. A free text search capability for title words is also provided in most systems.

Although the automated systems offer several advantages over manual searching of abstracts and indexes, they still provide a number of stumbling blocks to the identification of documents containing relevant information. Our experience with the four microbiologists has shown that questions representing a knowledge gap in a researcher's schema may be specific, naming an organism or a process of interest; may be general, looking for elements or members of a more general category

(e.g., exhaustive lists of members of a set); may be results oriented; or may be methodological. Existing systems are well-structured to address the specific type of question, however, they do not operate effectively in identifying members of a class or in selecting on methodological dimensions.

In responding to questions which are asking for information about members or elements of a set, the system must be able to link titles and abstracts containing element names with the name of the set. This requires that titles and abstracts be indexed at adjacent levels in a subject hierarchy. Thus, a question asking for the names of animal models would be responded to by titles that contained names of specific models such as 'rat reflux' or 'burned mouse'. There are several questions of this type in the present study. Some examples are listed below:

- What animal models have been used to study the effects of secondary pathogens?
- What extracellular products are produced by aeromonas hydrophila?
- What gram-negative bacteria metabolize ornithine?

The on-line bibliographic databases used to identify titles in response to these questions were not effective. The vocabularies of these systems were hierarchical but titles and abstracts were generally indexed at only the most specific level. As a result, the titles retrieved in response to the question on extra-cellular products contained the words 'extra-cellular products' but did not contain specific extra-cellular products such as endotoxin, hemolysin, elastase, etc. Indexing at multiple levels in a subject hierarchy may be an extremely complex process in a large multi-subject database, however, adjacent-category indexing of the database should be feasible in a more narrowly-defined content area such as vaccine development.

Responding to methodological as well as content (result-oriented) questions requires that titles and abstracts be indexed on methodological as well as content dimensions. There were several questions about methods which could not be easily answered with current indexing limitations. One example of this type of question is: What isolation methods have been used to recover streptococci from an aquatic environment? Existing systems were totally ineffective in responding to this question. Titles identified from the search did not even contain the words 'isolation methods'.

C. System Requirements for Analysis of Content

Titles resulting from a bibliographic search and lines of text in selected articles must be evaluated for relevance to questions representing the knowledge gaps in the researcher's schema. The evaluation and selection process required the system to make a series of content matches between questions and titles and between questions and the text. In performing these evaluations the system applied its knowledge of both the questions and the context-for-answers or schema representation for each question. The list of system tasks involved in content analysis include:

- Recognize usages, relationships, intents, actions, etc., represented in both the question and the potential response(s).
- Recognize definitions and synonyms as alternative ways of expressing key elements of a question.
- Recognize members of a general set identified in the question.
- Recognize the elements of the content model which define the context-for-answers.
- Recognize key elements in developing an adequate description of an experimental sample, or an experimental method.
- Recognize research results.
- Recognize type of study (e.g.: experimental, theoretical, etc.)

The system requirements in performing each of these tasks are described in the following sections.

1. Recognize Usages and Relationships

Key terms in a question, a title, or a section of text can be used in a variety of ways. A term can act, interact, effect, be acted upon, be compared with, etc. The system, through the use of syntactical rules, must be able to recognize that the question is asking about a specific usage of x or a specific relationships between x and y and that this desired usage or relationship is expressed in the title or text. An aid to the system in performing this task is a list of all the ways a specific term could be used. For example, the enzyme product exotoxin A can be used in the following ways:

- It is produced.
- It is treated.
- It has antibodies formed against it.
- It is compared with other toxins.
- It performs a role in disease development.
- It has a mode of action.
- It interacts with other toxins.
- It is diagnosed.
- It is studied, tested, assayed.

It is possible that all of the above descriptions might appear somewhere in the list of titles resulting from a search on exotoxin A. The system must be able to select those matching the description/usage called for in the question.

2. Recognize Definitions and Synonyms

The content of a title or a line of text can match the 'meaning' of a question without including the same terms. The professional literature is filled with alternative names for the same entity, process, procedure, and the system must be able to recognize these alternative names as meaning the same thing as key words in the question. If a researcher is

interested in learning about animal models used to study virulence factors in opportunistic pathogens the system must know that:

- Virulence factors are also pathogenic, toxic, disease causing, disease correlates, extracellular products, extracellular enzymes, etc.
- Animal models are also animals/hosts which are challenged, traumatized, altered, experimentally-infected, compromised, etc.
- Methods are also procedures, tests, assays, treatments, media, etc.
- Opportunistic pathogens are also secondary pathogens.

3. Recognize Members of A Set

Titles and lines of text may include members of a set specified in the question or they may include a more general category of which the set is a member. In the present application, many of the knowledge needs expressed by researchers were concerned with identifying members of a set. For example:

- Which bacteria metabolize ferric pyrophosphate?
- Which methods have been used to study ornithine metabolism by bacteria?
- Which isolation methods have been used to recover streptococci from the aquatic environment?
- What are the primary diseases of humans who die from pseudomonas infection?

In order to effectively address these types of questions the system must have hierarchical subject lists. It must have a list of the specific names for bacteria and know which are enteric and which are gram-negative; it must have a list of analysis methods like mass spectrum, protein decomposition, and isolation methods like membrane filter, most probable number; and it must know disease names such as cancer, pneumonia, wound, etc. Some examples of hierarchical subject lists needed by the system in creating products for the four microbiology researchers were as follows:

- Gram-negative rods.
- Gram-positive cocci.
- Enteric bacteria.
- Pathogenic roles.
- Recovery media for x.
- Animal models used in research on x.
- Measurable characteristics of x.

In cases where subject hierarchies are incomplete, the system must have some mechanisms or rules for inferring that a specific entity, process, or product belongs to a particular set. One rule is that an unknown element described with a series of known elements is probably a member of the set representing the known elements. For example, unknown elements appearing in tables with known elements can be assumed to be members of the same class. A sequence of words describing an organism can be considered as characteristics of that organism.

4. Recognize Elements of Content Models

The system needs to be able to recognize terms, statements, relationships, descriptions in the literature which are relevant to the researcher's need but do not directly match the terms or the synonyms for the terms in the question. Recognition of what is relevant is assisted by a knowledge of the content model that is being addressed by the question. The content model provides the system with a context for recognizing questions which are related to the question that was asked. Work with microbiologists in vaccine development led to the identification of several content models. One of particular interest was the content model for Virulence Factors (see Table 6, page 23). One researcher who asked about virulence factors had the following question: What role does each product of P. aeruginosa play in disease development singly and in combination? The model provided the system with knowledge of the several roles, lists of potential enzyme products, the fact

that modes of action may differ and the fact that various strains within the species may be more or less virulent. All of this information provided the system with a frame for searching the literature and identifying relevant statements. That is, the system knowing the content model, could match up the name of specific roles and products, and could search for descriptions of "modes of action" and comparisons between species.

5. Recognize Elements of Procedural Scripts

Researchers working on experimental problems usually have knowledge gaps which are addressed by information extracted or paraphrased from the experimental literature. Many of the questions asked in the present study required brief descriptions of the experimental sample and the experimental method as part of the answer. The format of articles reporting on experimental work includes separate sections on sampling and methods. In order to know or understand the composition of the sample or the key steps in the method, the researcher needs only selected statements not the entire description presented in the text. Effective selection of extracts or preparation of summaries required that the system be able to sort out those statements which are necessary for understanding. An aid to this sort and select step is a procedural script which specifies the key elements in sample selection and the key steps in conducting experiments. For example sample selection involves:

- Determining overall sample size.
- Determining if the sample will be divided into groups or subgroups based on some characteristics or set of characteristics.
- Determining the conditions under which the sample will be stored.
- Determining the original source of the sample (.e.g.: water, land, hospital, clinical specimen, etc.).

In creating a brief description of the experimental sample each of the above elements should be included. This can be accomplished by matching the elements of the script against the statements in the text.

6. Recognize Results, Categories of Study, Etc.

Many questions are concerned with findings or results. The researcher wants to know what has been found about a particular organism, process, product. In constructing answers, the system needs to be able to recognize a result. Result recognition can be guided and assisted by a list which specifies the various ways in which results are presented. For example, results may be in the following forms:

- Tables, figures, charts.
- Numbers in the text resulting from statistical computation.
- Statements about comparisons.
- Statements about significance.

Other capabilities of this sort are needed to enable the system to recognize the types of studies that are conducted (experimental, theoretical, review) and the author's rationalé for doing the study (e.g.: study effects, make comparisons, examine experimental treatments, etc.).

D. System Requirements for Responding to Interdependent Questions

Questions are interdependent when there is overlap in the material they cover. Question interdependence was encountered in the present study with the two researchers who asked multiple questions about one organism. This condition of content overlap led to a situation in which combinations of terms selected for one question resulted in citations which matched with other related questions. In order to effectively handle this situation the system should have the capability of evaluating every title against every question. An example of question relatedness is shown in Table 9. Here the search was organized around various aspects of the organism pseudomonas aeruginosa.

- Its toxic products: Exotoxin A, protease, elastase, hemolysin.
- How it is studied: animal model.
- Its role in disease: death by infection.

The questions or topics covered by these searches overlapped: each of the toxic products can be studied using animal models or could be involved in infections causing death. Animals used as models for study could die from infection caused by one ore more toxic products.

Table 9. Search Results for Pseudomonas Aeruginosa

Search Terms	Titles Identified	Titles Selected by System			
		Exotoxin A	Protease, Etc.	Animal Model	Death by Infection
Exotoxin A	23	15		3	
Protease, Elastase, Hemolysin	20		12	3	1
Animal Model	14			6	3
Death by Infection	55	1		6	11
Total	122	16	12	18	15
		61			

Experience in the present project has also shown that a journal article is likely to contain information that is relevant to a number of questions in addition to the question it was selected to answer. In the case where the researcher asks multiple, overlapping questions, the ideal system would have the capability to review each article as a potential source of relevant information for each question.

E. Summary

Search procedure and software requirements for conducting successful searches in a future schema/frame-based knowledge delivery system will depend on how scientific knowledge will be stored. Full-text document storage is a possibility; however, some extension of the work of Sager (1977) appears more efficient and also more compatible with schema/frame concepts of information structures in restricted domains of knowledge. Sager recognized that researchers in a subfield of science speak a specialized language which has grammatical regularities more restrictive than natural discourse. Using a subfield of pharmacology research as a context, she has developed a battery of computer programs which transform journal text into table-like information structures which carry the content of a sentence or sentence string. The columns of the tables are defined so as to preserve the syntactical relations between the words of the sentence; i.e., a paraphrase of the original text could be reconstructed from the tables. The advantage of the tabular format is to make the information content accessible for further computer processing. For example, files of articles on a common topic could be queried by computer for compilations of information with respect to particular categories.

We view the direction of Sager's (1977) work as organizing databases in subfields of science into frame-like information structures which potentially match the question schemata characterizing researcher information needs. Our experiences in conducting searches in the presently-configured database of citations and in screening full-text documents for answers relevant to researcher's question schemata are summarized as a source of system requirements in the direction of this goal.

The system was able to successfully search the professional literature and identify pieces of information that were relevant to the knowledge gaps in the researchers' schemata. Strong

guidance in this process was provided by the content and procedural models which surrounded the questions.

The content models provided a hierarchical context for selecting answers from the literature; they guided the system in identifying questions that were related to or hierarchically subordinate to the asked question. These 'new' questions served as an expanded frame for searching the literature.

The procedural models were used as aids to the selection of key descriptive information about sample selection and experimental procedure.

Other knowledge required by the system in performing effective search included:

- The variety of ways in which key terms can be used and related to each other.
- Definitions and synonyms for key words used in questions and in the context-for-answers surrounding those questions.
- The content hierarchies within which the researcher's need is embedded.
- Guides in the form of cues for recognizing research findings, study type and author rationalé.

VI. SYSTEM REQUIREMENTS FOR PRODUCT DESIGN AND PRESENTATION

A. System Procedure for Product Development

The simulation of a future, automated knowledge delivery system included a product design and evaluation phase. In our concept of this phase, the system was to integrate research findings over the set of informational materials identified during Search, and present as output, the researcher's old schema with the gaps filled in. The system was only marginally successful in achieving this ideal. At certain levels of question, the required answer was a list of items (e.g., animal models, enzyme products, methods of isolation, recovery media, etc.) and in these instances the system could compile from multiple sources, items which fit the list requirement. The system could also manage matrix products, or lists within lists, e.g., a list of appropriate media according to selected methods of recovery. At the level of integrated experimental results, however, we were unable to compile results on common frameworks. Original experimental works varied on too many dimensions for us to manage within the limits of our judgment of a reasonable level of effort. 'Animal model' research, for example, included variations in species of animal, type and extent of trauma, various characteristics of injected organisms or enzymes, time intervals between trauma and injection, organs examined for damage, measure or index of damage, etc. At the level of experimental results, the system regressed to preparing, in various ways, products which described single journal articles according to the perspective of the system's understanding of the researchers' question schema. From these experiences with product design and presentation we provide general findings about system requirements which enable the researcher to choose and direct the sequence of presentation and the format of the information to be presented.

B. Requirements for Researcher-Controlled Output

The system has diagnosed needs, searched the literature and pulled together relevant material, now how does the researcher want to learn about it? Our simulations suggest that a future knowledge delivery system will need to provide for user control of its output in at least two ways: the order of topical exploration, and the format characteristics of displayed products.

1. User-Directed Sequence

Researcher-system interactions in Diagnosis enabled the system to recognize relevant material and gather it together according to Question Sets in Search. The assimilation of the new material appeared to be made easy by enabling the researcher to direct the order of presentation. User control over presentation had additional advantages: it compensated for the system's lack of specificity about what the researcher already knew, and it enabled the researcher to follow-up new questions that arose in the presented material; i.e., the researcher recognized questions he did not know he needed to ask during Diagnosis.

For example, the system gathered information about Methods for Isolating Streptococcus from Aquatic Environments. A portion of the user-controlled presentation proceeded roughly as follows:

R: What can you tell me about Methods other than the Membrane Filter Test?

S: System displays a list of 'other methods'.

- Most Probable Number
- Pour Plate
- Precipitin
- Coagglutination
- Fluorescent Antibody

R: I forgot to mention it before, but I know a lot about MPN and Pour Plate. The other three are new to me; let's start with Precipitin.

S: System displays citations and article summaries.

R: These articles are all clinical or theoretical, no evidence of their use in aquatic applications. Let me look at Coagglutinate and F.A.

S: System displays citations and article summaries.

R: Article 2 looks promising; the medium is commercially available and the method has been used in natural waters, what were the findings?

S: System displays tables about organisms recovered, false positives, water temperatures and salinities.

R: This table suggests that recovery rates with this method may be sensitive to salinity level, I need to learn about the salinity variations in Chesapeake Bay.

2. User-Directed Formats

In the majority of instances, information products provided to the researchers were based on material taken from single documents, usually journal articles. Information from the article was selected by the system as relevant to the researcher's question schemata; the way in which selected information was presented to the researcher varied along several dimensions as follows:

- Use of Headings vs Unstructured Text
- Lists vs Discourse within Headings
- Excerpts vs Paraphrase
- Text vs Graphics

These variables were not studied experimentally; i.e, their variation was not systematically controlled, but were iteratively varied to conform to a given user's preference. These particular system interactions in information presentation developed three types of product designs: document characterization, document synopsis-with-pointers, and document condensation.

a. Document Characterization. Document characterizations best served the purposes of the researcher who was at the stage of transitioning from learning how to isolate selected genera of organisms to initiating an experimental program of virulence factors determination. He wanted information at the level of a survey of the experimental field with respect to these genera; information at the level of what was being studied rather than any details about methods or results. The product design that suited his needs was a characterization; the product described what the article was about. Categories of information in the characterization included type of article, organisms studied, methods used, phenomena studied, type of result reported. Where possible lists were used rather than discourse and shorter characterizations were preferred to longer. All characterizations were less than a page in length. A typical product of this type is illustrated by the characterization in Figure 5.

b. Document Synopsis-with-Pointers. This class of product design was in two parts: first, a brief summary or overview of the article which contained parenthetical references (pointers) to sections or tables in the original document which amplified the summary statement; second, a set of excerpts from the original document which contained the elaborations pointed to in the overview. The system developed this kind of product with two of the researchers whose content interests were most difficult for the system to comprehend. The synopsis-with-pointers product design enabled the researcher to judge the relevance of the work from the synopsis, then selectively explore details of interest by calling up the appropriate pointed-to material. Also, this design strategy enabled the system to excerpt blocks of information it only weakly understood, but could recognize as relevant.

The synopses were very short, three to eight sentences in an unstructured format, i.e., there were no headings, but each included major topics of purpose, organism, method, result. Figure 6 illustrates typical synopses of journal articles about the Most Probable Number Method for isolating gram-positive cocci from aquatic environments. Each 'pointer' in the synopsis called up an excerpt from the original text which was reproduced on a single page and presented to the researcher on demand. This product design provided a high degree of flexibility for the researcher, enabling him to selectively scan the article to whatever level of detail he required and in whatever order of topic.

c. Document Condensation. These were highly-formatted, highly focussed, full-document substitutes. They included one or two sentences per heading and the headings paralleled journal style with the addition of two headings which the researcher preferred: Kind of Study and Author's Rationalé. A fully described article included seven headings as follows:

- Kind of study
- Author's rationalé
- Objectives/purposes
- Organisms and phenomena studied
- Methods of assay
- Results
- Author's interpretations/conclusions

Condensations were 1-3 pages in length which represented a significant compression of original articles. With the exception of Results, content of the condensations were system-generated descriptions which summarized and paraphrased the author's work within the main headings. The product design for the Results heading was for the system to generate a result statement and excerpt from the original document one or more graphics which supported the result statement. A sample condensation is shown in Figure 7.

The document condensation suited the needs of the researcher who was most advanced in the vaccine development paradigm, and needed detailed information from related experimental work against which to compare his results. Of the three product types, the preparation of article condensations placed the most demanding requirements on the system, i.e, required the strongest procedural and content models. They were highly satisfying to the user, however. They were relevant, highly focussed, and their content could be absorbed in a few minutes time. Given the condensation, the researcher expressed no need for the original document.

C. Summary

A future knowledge delivery system will need to provide for user control of its output; enabling the researcher to direct the presentation sequence and select from among alternative presentation formats.

Control over presentation sequence enables the researcher to pursue topic areas in a preferred order and to desired levels of detail.

Control over format allows the researcher to chose a mode of presentation (e.g., summaries, graphics, text excerpts) which best meets his needs. There is a wide range of potential format combinations that could be made available to a researcher; the current project made effective use of three: Document Characterization, Synopsis-with-Pointers, and Document Condensation.

Evans, N. Pathogenic mechanisms in bacterial diarrhoea. Clinics in Gastroenterol., 1979, 8(3), 599-623.

This is a review article. It includes one small section on campylobacter jejuni which discusses its occurrence as an intestinal pathogen and the methods for isolation. Suggested mechanisms for causing diarrhoea are tissue invasion or cytotoxin.

Other agents of diarrhoea presented include: enteropathogenic E. coli, enterotoxigenic E. coli, enteroinvasive E. coli, cytotoxic E. coli, salmonella, shigella, vibrios, food poisoning bacteria, clostidia and bacteria associated with enterocolitis.

Discussion also covers bacterial plasmids of E. coli and pathogenic mechanisms as follows:

- Enterotoxins
- Choleragen
- Heat-labile (LT) and heat stable (ST) toxins of E. coli
- Ent plasmids
- Cytotoxins
- Bacterial adhesion (mainly E. coli, salmonella and shigella)
- Bacterial invasion

Also included are host and environmental factors and implications for treatment and prevention.

Figure 5. Document Characterization

Phase 2: Strain Selection

Task E: Outline procedures for isolating desired strains

Question Set: What isolation methods have been used in aquatic environments?

Specific Question: What methods other than Membrane Filter have been used to isolate gram positive cocci from an aquatic environment?

A. MPN - Most Probable Number

1. (Mallman, W. L., and Seligmann, E. B., 1950)

   Compares four media for detecting streptococci in water: standard lactose broth, sodium azide broth, SF broth and azide dextrose borth (p. 287). The most probable number method was used according to Hoskins' tables. Also, microscopic determinations were made. Tables 1 and 2 (p. 288) compare media in river water and in swimming pools. Azide dextrose is the best media, however, samples must be checked microscopically. SF broth gave the lowest indices.

2. (Litsky, W., et al, 1953)

   Proposes presumptive and confirmatory media for enterococci. Azide dextrose broth (p. 876) and ethyl violet azide broth (p. 877). Ethyl violet was added to remove gram-positive bacteria other than enterococci. Table 1 (p. 878) compares SF broth, a media prepared by Winter and Sandholzer (1946) and dextrose azide and ethyl violet azide broth. The DA and EVA broth were superior by 100-1000 percent. The MPN method was used.

3. (Kinner, B. A., et al, 1960)

   KF media was developed and tested against BAGG broth and DA-EVA. KF media was superior (p. 18, Table 3). Pour plate and MF also discussed and compared. Media composition shown (p. 16, Table 1).

Figure 6. Document Synopses-with-Pointers

Al-Dujaili, A. H., and Harris, D. M. Pseudomonas-aeruginosa infection in hospital: A comparison between infective and environmental strains. J. Hyg., 1975, 75(2), 195-201.

Kind of Study

This is an assay of various strains of P. aeruginosa taken from hospital patients and from the general hospital environment.

Study Rationalé

The idea of the study is that strains of P. aeruginosa vary in virulence, which explains the sporadic nature of hospital infection Study purpose was to assay the products of strains assumed to be infective and non-infective in order to discover differences.

Strains Isolated

Fifteen strains of P. aeruginosa were isolated and tested for production of extra cellular toxins. Samples were taken of 156 patients with P. aeruginosa infection and from various environmental sites in hospital: sinks, mops, baths.

Methods of Assay

Methods of Liu, Abe and Bates, 1961, were used for separating factions Ia and Ib (pyocyanin and other pigments) faction II (haemolysin) and faction III (protease, lecithinase and lipase).

Results

Results suggest that high haemolytic titre may be the virulence factor. (See Table 3)

Table 3. Biological activity of fractions II and III of selected 'infective' and 'environmental' strains of Ps. aeruginosa

Strain No.	Pyocine; serotype	Source	Ass. infections	Reciprocal of titres: Haemo-lysin	Protease	Leci-thinase	Lipase
1	3 ; 6	Patient	8	16	32	4	2
24	3 ; 6	Patient	8	32	32	16	16
23	3 ; 6	Patient	8	32	16	4	2
32	1(c); 3	Patient	7	16	16	8	< 2
61	1(c); 8	Patient	7	16	8	2	2
150	1(a); 5	Patient	2	16	8	16	< 2
158	10 ; 11	Patient	5	32	4	16	< 2
58	1(a); 7	Patient	2	8	8	4	4
157	uc ; 6	Patient	6	8	8	8	2
17	10 ; 13	Environment	1	2	16	16	8
26	1(g); 6	Environment	0	4	4	2	2
15	uc ; na	Environment	0	2	32	8	2
171	34 ; 3	Environment	0	2	8	8	2

Figure 7. Document Condensation

VII. REFERENCES

Bartlett, F. C. Remembering: A study in experimental and social psychology. Cambridge, Gr. Brit.: Cambridge University Press, 1932.

Brown, J. S., and Burton, R. R. Multiple representations of knowledge for tutorial reasoning. In D. Bobrow and A. Collins (Eds.) Representation and understanding: Studies in cognitive science. New York: Academic Press, 1975.

Hoover, R. E. The library and information manager's guide to on-line services. White Plains, N.Y.: Knowledge Industry Publications, Inc., 1980.

Kennedy, J. L. The system approach: Organizational development. Human Factors, 1962, 4, 25-52.

Kuipers, B. J. A frame for frames. In D. Bobrow and A. Collins (Eds.) Representation and understanding: Studies in cognitive science. New York: Academic Press, 1975.

Mavor, A. S., and Vaughan, W. S., Jr. An assessment of mechanisms for international exchange of science information. Annapolis, Md.: W/V Associates, November 1980.

Minsky, M. A framework for representing knowledge. In P. H. Winston (Ed.) The psychology of computer vision. New York: McGraw-Hill, 1975.

Norman, D. A., and Bobrow, D. G. On the role of active memory processes in perception and cognition. In C. N. Cofer (Ed.) The structure of human memory. New York: Freeman and Co., 1975.

Parsons, H. Mc. Man-machine system experiments. Baltimore: The Johns Hopkins Press, 1972.

Rumelhart, D. E. Notes on a schema for stories. In D. Bobrow and A. Collins (Eds.) Representation and understanding: Studies in cognitive science. New York: Academic Press, 1975.

Rumelhart, D. E., and Ortony, A. The representation of knowledge in memory. In R. C. Anderson, R. J. Spiro, and W. E. Montague (Eds.) Schooling and the acquisition of knowledge. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Pub., 1977.

Sager, N. Information structures in the language of science. In E. C. Weiss (Ed.) The many faces of information science. Boulder, Colo.: Westview Press, 1977.

Schank, R. C. The role of memory in language processing. In C. N. Cofer (Ed.) The structure of human memory. San Francisco: W. H. Freeman and Co., 1975a.

Schank, R. C. The structure of episodes in memory. In D. Bobrow and A. Collins (Eds.) Representation and understanding. New York: Academic Press, 1975b.

Schank, R. C., and Abelson, R. Scripts, plans, goals and understanding. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Simon, H. A. The sciences of the artificial. Cambridge: MIT Press, 1969.

Spiro, R. J. Remembering information from text: The state-of-schema approach. In R. C. Anderson, R. J. Spiro and W. E. Montague (Eds.) Schooling and the acquisition of knowledge. Hillsdale, N.J.: LaWRENCE Erlbaum Associates, Pub., 1977.

Weiss, E. C. (Ed.) The many faces of information science. Boulder, Colo.: Westview Press, 1977.

Winograd, T. Frame representations and the declarative/procedure controversy. In D. Bobrow and A. Collins (Eds.) Representation and understanding: Studies in cognitive science. New York: Academic Press, 1975.

APPENDIX A

PROCEDURE FOR GENERATING A CONTEXT FOR QUESTIONS

Interview Procedure

Step 1: Identify Researcher's Topic and Methodological Paradigm

A. Interview Aids

1. List of research paradigms
2. List of phases for each paradigm (list or flow chart)

B. Questions

1. What is the topic/problem area of your research?
2. This is a list of the research paradigms in microbiology. Which research paradigm/approach are you using to study this problem?
   - Show researcher the list of research paradigms.
3. This is a model of the paradigm you selected. Do the phases of your research match the phases shown in this list/flow chart?
   - Show researcher a flow chart/list of the phases of the selected paradigm.
4. If your process or planned process does not follow this model, how would you change the model to fit your work? Add, eliminate, etc.? Please discuss the specific changes.

C. Products Created by the System

1. A research topic.
2. A methodological paradigm which matches the researcher's process.

Step 2: Determine Location of Researcher in Selected Paradigm

A. Interview Aids

1. A list or flow chart of the research phases of the selected paradigm.

B. Questions

1. What phase of your project are you currently working on? Show the paradigm as an aid for the researcher in indicating where he is in the process.

C. Products Created by the System

1. Current phase and task area or areas.

Step 3: Describe Project Decisions that Have Already Been Made

A. Interview Aids

1. Research phases of the selected paradigm.
2. A task/activity breakdown or listing for each completed phase.

B. Questions

1. Specific questions here will depend on both the selected paradigm and the phases that have been completed in that paradigm.
2. The researcher will be asked to describe decisions, activities, topics avoided, ideas explored and discarded for each completed phase. Show the reseacher the list of tasks associated with each phase and ask him to comment about the decision/activity/outcome.

C. Products Created by the System

1. An overview of completed phases or task areas composed of:
   - Decisions
   - Methods used
   - Results obtained
   - Topics deliberately avoided
   - Ideas already explored and discarded

These descriptions will be generated for each completed phase of the research. The descriptions give a context to both the researcher and the system. This is where you've been – where do you go from here, what are the problems now?

Step 4: Identify Current Problem and Partition It into Small, Manageable Units

A. Interview Aids

1. Outline of tasks for current research phase.

B. Questions

1. Here is a list of tasks associated with the current phase of your research project. Do these tasks match with your idea about the activities/ decisions that occur in this phase? Show the list of tasks.
2. If not, what tasks would you change, eliminate? Are there tasks that have not been included that should be listed?
   - Changes should be made by the system to reflect the researcher's response.
3. Which of these tasks have you completed? What were the outcomes of these tasks?
4. What are you doing now on task ___ (the next task) and what is the problem you are experiencing?
5. What are the specific pieces of the problem? These may relate to steps within tasks.
   - Try to break problem into small manageable units.

C. Products Created by the System

1. A list of the tasks for the current research phase as seen by the researcher.
2. Decisions/activities/outputs of tasks completed or partially completed.
3. The current problem broken down into small, manageable units.

Step 5: Structure and Describe Researcher's Knowledge State for Each Unit of the Problem

A. Interview Aids

1. A list of problem units for current task(s).
2. A frame for describing the researcher's existing knowledge state.

B. Questions*

1. Here is a form which shows the kind of information needed to describe your knowledge needs. We will start with Task ____ and problem ____.
   - Fill in task and problem
   - Show researcher the form
2. What are the facts you have about this problem? What are the sources of these facts? Literature, colleagues, observation, etc.?
3. What are the inferences you have made about this problem? What are the bases of these inferences? Are they based on analogy?
4. What are the specific pieces of knowledge you need to proceed to the next part of the problem, the next step, the next task?

C. Products Created by the System

1. A knowledge schema for each problem unit which includes facts, inferences, and gaps.

*The same procedure and set of questions will be used to describe the researcher's knowledge state for each problem unit.

Step 6: Determine the Characteristics of the Information That Will Fill in the Knowledge Gap(s)

A. Interview Aids

1. Knowledge schema for each problem unit constructed in Step 5.
2. Categories of answers.

B. Questions

1. Here is a list of the types of information you could be seeking to substantiate your inferences and to fill in gaps in your existing knowledge. Which category best fills each of your needs?
   - Show list of categories
   - Needs should be taken in order.

C. Products Created by the System

1. The form of the information which best matches each knowledge need.

Note: It may be best to do steps 5 and 6 together for each knowledge schema. That is, complete one and then start the next.

APPENDIX B
CASE STUDIES

Dr. Olgerts R. Pavlovskis
Medical Microbiology Branch
Naval Medical Research
Institute
Bethesda, Maryland

I. DIAGNOSIS: CONTEXT AND QUESTIONS

A. Context for Identifying Questions

1. Statement of the research project: Experimental examination of pseudomonas aeruginosa virulence factors and their role in disease.

2. Statement of the topic's importance; a justification of the research:

   a. During Viet Nam war, significant numbers of military personnel died of septic shock following burns, wounds and amputations (sepsis is evidence of pseudomonas activity).

   b. During peacetime, there is an everage of one burn case per day among boiler-room personnel aboard Navy ships.

   c. Pseudomonas infection is difficult to treat, i.e., has a high level of resistance to antibiotics, so preventive vaccine is a priority need.

3. Statement of the goal structure: to produce vaccine against pseudomonas aeruginosa.

4. Characterization of the researcher's experience/level of sophistication with the topic

   The researcher has completed a series of experimental studies on toxic products of p. aeruginosa using a burned mouse animal model.

5. Characterization of the status of research on the topic:

   a. It is an opportunistic or secondary pathogen; harmless to healthy persons/animals, but pathogenic to traumatized patients.

b. It produces a variety of enzymes some of which have known pathogenic roles:

- Protease - enables organism to establish itself in hosts.
- Exotoxin A - enteres the healthy cell and prevents protein synthesis. Acts like diphtheria toxin.
- Elastase, hemolysin, endotoxin - other enzymes whose roles are not well-known.

c. Three approaches, lines of attack against p. aeruginosa, are being explored in vaccine development.

(1) Attack the organism directly, destroy the cell.

(2) Attack the organism's pili which are instrumental to attachment in the host.

(3) Neutralize the organism's toxic products.

d. 'Compromised Host' model is used for experimental work in this field. Since p. aeruginosa is only pathogenic to traumatized patients, experiments are conducted with animals traumatized in some way. His laboratory uses a "Burned Mouse Model" for experimentation. Various antibodies are tested for effectiveness against p. aeruginosa bacteria injected into a burned mouse.

6. Statement of excluded topics:

a. Researcher is interested in one strain of bacteria only, i.e., pseudomonas aeruginosa, and no others.

b. Researcher is interested in the series of links from bacteria to toxic products to bacteremia to organ failure and death in compromised hosts; not in the genetics, characteristics or biochemistry of p. aeruginosa.

7. Characterization of the researcher's current problem:

The researcher is at the stage of experimental determination of actions and interactions of p. aeruginosa toxic products as they affect host cells. He has worked extensively with one animal model and is looking for others to use. He also wanted to link his results to the causes of death in humans.

B. Questions

1. Question Set #1: Virulence Factors of P. Aeruginosa

Asked Question 1. What role does each product of p. aeruginosa play in disease development, singly and in various combinations?

Unasked Question 1-a. What is the mechanism by which a given enzyme (product) accomplishes its role/function in disease development?

Unasked Question 1-b. What differences have been identified in enzyme product characteristics between virulent and non-virulent strains of p. aeruginosa?

The researcher asked for information about roles of p. aeruginosa enzyme products in disease development. He gave the following two examples: protease enables the bacteria to establish itself in the host, and exotoxin A destroys the host cell by inhibiting protein synthesis. The mode of action of exotoxin A is analogous to that of diphtheria toxin.

The system made the inference from researcher's example about exotoxin A that he was also interested in mechanisms or modes of action, so the system generated Question 1-a from the interaction.

During the process of scanning the retrieved journal articles for answers to questions 1 and 1-a, the system recognized a series of experiments in which enzyme products of virulent and non-virulent strains of p. aeruginosa were compared/contrasted. The system recognized these studies as relevant to the researcher's interest since they were attempts to determine those characteristics of p. aeruginosa enzymes which were disease correlates. So the system generated Question 1-b during the literature review procedure.

2. Question Set #2: Death from P. Aeruginosa Infection

Asked Question 1. What is the cause of death in patients who die of septic shock attributable to P. aeruginosa?

Asked Question 1-a. What organs are damaged by P. aeruginosa (e.g., heart, liver, lung)?

Asked Question 1-b. Where in the victim's body were P. aeruginosa found (e.g., respiratory system, renal system, wounds)?

Unasked Question 1-c. What were the principal illnesses (diagnostic categories) of the bacteremic patients who died vs those who survived?

Unasked Question 1-d. What is the frequency of occurrence of bacteremia in hospitalized patients?

Unasked Question 1-d(1). What is the mortality rate among bacteremic patients?

Unasked Question 1-d(1)(a). What are the historical trends in both frequency of occurrence and mortality rate?

The researcher indicated a lack of knowledge about the ultimate cause of death in hospital patients where septic shock from P. aeruginosa infection was a contributing cause. He asked three questions. one general and two more specific questions which elaborate the general question. He directed us to look for articles where autopsies had been performed.

Several information products were prepared by the system which provided information relevant to these three questions, but the researcher needed more than we gave him. He wanted to know about the primary illness of the victims and about statistical trends in frequency of occurrence and mortality rates. The latter topics helped to document the importance of the researcher's work, justified research on the problem of vaccine development.

3. Question Set #3: Animal Models Used in Experimental Research with Opportunistic Pathogens

Asked Question 1. What animal models are being used in research on experimental infection by opportunistic pathogens?

Unasked Question 1-a. What are the procedures required to implement each animal mode:

Unasked Question 1-a(1). What are the equipment and instrumentation requirements of each animal model?

The researcher had instructed the system about the 'animal model' paradigm for experimental research with opportunistic pathogens. His own laboratory used a 'Burned-Mouse' model.

He expressed a need for information about the various animal models which were in use (Question 1). The system identified some citations, retrieved articles and prepared information products at the level of 'what models'. The system then learned that the researcher wanted to know about the specific laboratory procedures (Question 1-a) and the instrumentation/equipment requirements (Question 1-a(1)) of the various models to that he could judge whether or not the identified models could be implemented in his laboratory.

C. Question-Specific Context (Knowledge Structures Required for Answer-Seeking)

1. Question Set #1: Virulence Factors

Question 1. About roles of P. aeruginosa products in disease development.

R a. P. aeruginosa is a bacteria that produces a variety of enzymes.

R b. Enzyme products of P. aeruginosa include the following:

- Exotoxin A
- Protease
- Elastase
- Hemolysin
- Endotoxin

R c. To bring about disease in a host, the bacteria must accomplish three functions (roles).

- Penetrate the host cell
- Attach itself to the host cell
- Damage the host cell

R d. There is evidence that exotoxin A destroys host cells by protein synthesis inhibition.

R e. There is evidence that protease facilitates either the penetration or the attachment function (role).

Question 1-a. About mechanisms or modes of action

SL* a. There are a variety of ways by which an enzyme product of P. aeruginosa can accomplish a given function.

SL b. The mechanism may vary with class of host cell under attack, i.e., lung tissue vs kidney tissue, etc.

Example: Hemolysin may contribute to lung cell penetration by reducing the effectiveness of alveolar macrophage. Alveolar macrophage attack bacterial cells. Hemolysin attacks alveolar macrophage. Alveoli are lung tissue cells.

R** c. The mode of action may be understood by analogy to the action of other bacteria produce toxins.

Example: Exotoxin A from P. aeruginosa destroys host cells by the same mode of action as does diphtheria toxin, i.e., by inhibiting protein synthesis.

Question 1-b. About enzyme differences between known virulent and non-virulent strains of P. aeruginosa.

R a. P. aeruginosa is the name of a genus and species of bacteria of which there are many strains.

R b. Some strains of P. aeruginosa are more virulent than others.

R c. Virulence is a function of enzyme characteristics and strains differ in the specific combination of enzymes they produce.

R d. P. aeruginosa is a very common bacteria, can be found most anywhere.

SL e. One method for determining enzyme correlates of virulence is to contrast the enzyme characteristics of bacteria taken from persons known to be suffering P. aeruginosa infection with bacteria taken from the general environment.

*SL - System Learning
**R - Researcher

SL f. Enzyme characteristics that can be measured and therefore used to contrast virulent vs non-virulent bacterial strains include:

- Pyocine type
- Concentration (titre)
- Amount

Some of the information in the above structures was provided by the researcher during the diagnostic interaction (R) other information was learned by the system during the process of reviewing the retrieved journal articles (SL).

2. Question Set #2: Death from Infection

Question 1. About cause of death in patients who died of septic shock attributable to P. aeruginosa

R a. P. aeruginosa is a secondary or opportunistic pathogen; harmless to healthy persons/animals but pathogenic to very ill or traumatized patients.

R b. Septic shock is a consequence of bacteria in the bloodstream.

SL c. The penetration/invasion of the blood by bacteria is designated by a variety of labels:

- Sepsis
- Bacteremia
- Septicemia

R d. P. Aeruginosa is a common bacteria, found everywhere, particularly prevalent in hospital environments.

SL e. P. aeruginosa must establish itself in the patient before it invades the bloodstream (e.g., in a wound or burn, in the respiratory system, urinary system, etc.).

SL f. The series of events leading to death with P. aeruginosa as a component is as follows:

- A very ill or traumatized patient in a hospital.
- P. Aeruginosa in the general hospital environment.
- P. aeruginosa established in the patient's wound, or system.
- P. aeruginosa penetration of the patient's blood stream (bacteremia)
- Eventual failure of a major organ (e.g., heart, liver, lungs)

Question 1-a. About organ failure

R a. The ultimate cause of death is the failure of a major organ; i.e., heart, liver, lungs, kidneys, etc.

R b. Major organs are examined and assessed as part of an autopsy.

Question 1-b. About source of P. aeruginosa in the patient's body

SL a. P. aeruginosa must first establish itself in the host's body prior to penetration of and colonization in the bloodstream.

SL b. Main possibilities are the respiratory system, renal system, wounds and burns.

Question 1-c. About main diagnostic categories of bacteremic patients

SL a. Hospitalized patients are described by diagnostic categories.

SL b. The route of P. aeruginosa colonization of hospitalized patients as well as the outcomes may vary systematically with the principal diagnostic category.

Question 1-d. About frequency of occurrence and mortality from P. aeruginosa bacteremia in hospitalized patients

SL a. Hospitals keep yearly records of bacteremia in patients so that historial trends can be plotted.

SL b. Increase in number of bacteremia occurrences and mortality rates is an index of the importance of this problem for research.

3. Question Set #3: Animal Models

Question 1. About animal models used at other laboratories

R a. Since P. aeruginosa is only pathogenic to already-ill persons, experimental research is conducted using animals which are made ill in some way, i.e., the animal is a simulation of an ill person.

SL b. An experimentally-made-ill animal is called an 'animal model'. There are a variety of synonyms for 'animal model'; these include:

- Compromised animal/host
- Challenged animal/host
- Altered animal/host
- Experimentally-infected animal/host

SL c. The research literature tends to label the animal model by the type of illness or trauma followed by the animal traumatized. For example:

- Burned-mouse model
- Reflux-challenged rat model

R d. P. aeruginosa is not the only opportunistic pathogen; others have generic names that include:

- Aeromonas
- Yersinia
- Campylobacter

Question 1-a. About laboratory procedures for implementing an animal model

a. The end-product of an animal model is an ill, burned or wounded animal; the specific technique by which the animal is weakened is of interest to the researcher.

b. Recent research literature will typically refer to an animal model by its name only, and refer to an older article for details of the procedure by which it is implemented.

Question 1-a(1). About the instrumentation, equipment and facility requirements of an animal model

a. Procedures for implementing a given animal model specify the facilities, equipments, and instruments involved.

II. SEARCH STATEGY AND IMPLEMENTATION

Two data bases were selected to address the questions relating to Dr. Pavlovskis' research with pseudomonas aeruginosa: BIOSIS and MEDLINE. BIOSIS was selected because it indexes the literature on bacteria; MEDLINE was chosen because it indexes the literature on the clinical effects or aspects of bacteria. Dr. Pavlovskis is interested in the toxic products (enzymes) of a bacteria (pseudomonas aeruginosa), the animal models used to study these enzymes, and the effects of these enzymes on humans.

A. Virulence Factors

The researcher named several toxic products (enzymes) of pseudomonas aeruginosa. They included:

- Exotoxin A
- Protease
- Elastase
- Hemolysin

There were three questions: what role do these products play in disease, how do these products work and what are the enzyme differences between virulent and non-virulent strains.

The first strategy was to combine all of the enzyme names with pseudomonas and aeruginosa.* BIOSIS included all of these index terms. The result was 185 citations. The system judged this set to be too large and chose to pursue exotoxin A separately. Exotoxin A was singled out because the researcher had said that it was of primary interest. The combination of exotoxin A and pseudomonas aeruginosa resulted in 86 citations. This set was also judged as too large; the

*In BIOSIS pseudomonas aeruginosa is not a term. It is necessary to enter each word as a separate term.

system elected to printout the most recent 20. In MEDLINE a similar procedure was used by linking the organism with exotoxin A. This led to nine citations. There were six citations which were common to both data bases; overall 23 unique titles were identified.

Of the 23 identified citations, 16 were selected by the system as being relevant to the researcher's questions. The system's objective was to be as targeted as possible - to select only those citations which directly address the researcher's questions. The seven citations that were not chosen were rejected because they appeared to deal with antibodies to the toxin or the genetics of the toxin rather than the toxins role, mode of action or the degree of virulence of selected strains. An example of a rejected title is, "Enzyme linked immuno sorbent assay for the measurement of mouse and human antibodies to pseudomonas-aeruginosa exotoxin A."

The sixteen selected citations fell into one of the following four categories:

- The role of exotoxin A in disease (4). All these titles employ the effects of toxin on the host or host cell.
  - "Roles of exotoxin and protease as possible virulence factors in experimental infections with pseudomonas aeruginosa"
  - "Toxicity of pseudomonas aeruginosa exotoxin A for human macrophages"
  - "Toxic products of pseudomonas aeruginosa production by isolates from cystic fibrosis patients and effects on human cells in-vitro"
  - "Experimental studies of the pathogenesis of infections due to pseudomonas aeruginosa"

- The structure or mode of action of exotoxin A (3). Some examples are:
  - "Structure-activity relationships in diphtheria toxin and exotoxin A from pseudomonas aeruginosa"
  - "Mechanims of action of pseudomonas aeruginosa exotoxin A in experimental mouse infections: adeno-sine diphosphate ribosylation of elongation factor 2"
  - "Enzymically active fragment of pseudomonas aeruginosa exotoxin A"
  - "Modes of action of diphtheria toxin and exotoxin A from pseudomonas aeruginosa"
- Comparisons of virulent and non-virulent strains (1)
  - "Production of exotoxin, protease and elastase of pseudomonas aeruginosa strains isolated from patients and environmental specimens"
- Methods of study - ways of examining exotoxin A. This was not a question asked by the researcher. It appeared to the system that methods of analysis (especially if new) would/could be of use to the researcher (3).
  - "Enzyme-linked immunosorbent assay for pseudomonas aeruginosa exotoxin A"
  - "Exotoxin A of pseudomonas aeruginosa, the secretion and isolation of membrane bound toxin"

The search for other citations on toxic enzymes was conducted using the following term combinations:

BIOSIS	MEDLINE
Pseudomonas and	Pseudomonas aeruginosa and
Aeruginosa and	Hemolysins or
Protease or	Peptide hydropases or
Elastase or	Pancreatopeptidase and
Hemo(w)lysin and	Disease(s)
Disease or	
Infection	

The BIOSIS search resulted in 17 citations (3 were on earlier printouts); the MEDLINE search provided 10 citations (6 were on earlier printouts). There was one article in common from the two searches. Thus, 17 new citations were identified; 14 of these were selected as relevant to the user's questions on toxic enzymes. The rejected articles were on antibodies, antitoxins and antigens. For example:

> "IgE antibody production to exoenzymes in common antigen (OEP) of pseudomonas aeruginosa in mice"

The selected titles were almost all directed toward the role of toxic enzymes in disease (9). Some examples follow:

> "The role of hemolysin in corneal infections with pseudomonas aeruginosa"
>
> "The effect of protease production by pseudomonas-aeruginosa on growth in burned mouse skin extract"
>
> "Toxic activity against alveolar macrophages of products of pseudomonas-aeruginosa isolated from respiratory and non-respiratory sites"

There was two citations on mode of action:

> "Protease and elastase of pseudomonas aeruginosa: In activation of human plasma alpha 1-proteinase inhibitor"
>
> "Study of mechanisms of pathogenicity of pseudomonas aeruginosa experimental infection of the mouse"

And one citation on strain comparisons:

> "Pseudomonas aeruginosa infection in hospital: A comparison between infective and environmental strains"

There were two other citations that were selected. One was chosen because it dealt with mouse burn infection and Dr. Pavlovskis is working with a burned mouse model.

> "Effects of somatic component of pseudomonas-aeruginosa on protective immunity in experimental mouse burn infection"

The other dealt generally with infection and a method of diagnosis:

"Pseudomonas aeruginosa infection and its sero diagnosis"

B. Death by Infection

The search terms used for BIOSIS and MEDLINE are listed below. These terms were selected to relfect the major concerns of the researcher: the cause of death in patients who die of septic shock attributable to pseudomonas aeruginosa, the organs damaged, the location of the bacteria, the principal illness of the patient and the frequency of occurrence in hospital patients.

BIOSIS	MEDLINE
Pseudomonas and	Pseudomonas aeruginosa and
Aeruginosa and	Death or
Infection and	Mortality or
Death or	Lethal or
Dead or	Fatal
Fatal or	
Kill or	
Lethal or	
Mortality	

Neither system had index terms for organ damage or mortality trends. The BIOSIS search terms identified 33 articles; the MEDLINE search terms identified 24. Discounting overlap there were approximately 55 new titles identified. Careful screening of the printouts led to the selection of 11 articles as being relevant to the researcher's questions on death, infection and pseudomonas aeruginosa. Six of the articles in this printout were more relevant to animal models and two were directly concerned with virulance factors. Those citations that were judged as not relevant fell into the following categories.

- The article appeared to be about prevention or treatment (e.g.):
  - "Pseudomonas aeruginosa: prevention better than cure"
  - "Treatment of patients with pseudomonas endo carditis with high does amino glycoside and carbonicillen therapy"
- The article appeared to be about methods of identification or purification (e.g.):
  - "Rapid identification of gram-negative rods from blood cultures using direct inoculation of the API-20C system"
  - "Purification and characterization of leucocidin from pseudomonas aeruginosa"
- The article appeared to be about the source of the bacteria (e.g.):
  - "Bacteremia: the significance of outside vs inside hospital origin"

The selected titles addressed four question areas: causes of death, organs damaged, frequency of occurrence in hospitals, and trends over several years.

- An example of titles related to causes of death is:
  - "Fatal bronchopneumonia and dermatitis caused by pseudomonas aeruginosa in an Atlantic bottle-nosed dolphir
- An example of titles related to organ damage is:*
  - Septicemia and bilary tract obstruction"
- An example in the frequency of occurrence category is:
  - "The incidence of pseudomonas aeruginosa infections from an open burn ward"
- An example in the trend category is:
  - "Bacteremia at Boston City Hospital, Mass, USA: Occurrence and mortality during 12 selected years: 1935-1972 with special reference to hospital acquired cases"

*System needs to know all things considered to be organs.

C. Animal Models

There were three questions on animal models: what animal models are being used in research on experimental infection by opportunistic pathogens; what procedures are used to implement each model; what are the equipment and instrumentation requirements of each model? The search was done at the level of animal models as the BIOSIS and MEDLINE data bases are not indexed to retrieve procedures and instrumentation. The search terms used in each data base are:

BIOSIS	MEDLINE
Pseudomonas or Aeromonas or Opportunistic (w) pathogen or Secondary (w) pathogen and Animal (f) model	Pseudomonas aeruginosa and Animal model

The BIOSIS search resulted in five citations. The initial MEDLINE search included combinations of secondary or opportunistic pathogens with animal models, however, this combination resulted in irrelevant citations (e.g.):

- Heat-stable somatic antigens of a group of unclassified fluorescent pseudomonads (UPF).

None of these referred to animal models or the use of animal models with secondary pathogens. For this reason, the final search only used pseudomonas aeruginosa and animal models as key words. This search identified nine citations. The total number of citations from both searches was 14; six of these were selected as relevant. In addition, citations on animal models resulted from previous searches on exotoxin A (3) and death by infection (6). The rejected citations

were more about death by infection than about the use of animal models. That is, they did not mention animal model or a specific animal. Some examples were:

> "Pseudomonas aeruginosa vasculitis and bacteremia following conjunctivitis: A simple model of fatal pseudomonas infection in neutropenia"
>
> "Antibiotic therapy and prophylaxis of experimental endo carditis"

Those articles selected as relevant included the words "Animal Model" or "Compromised Host" or "Altered Host" or "Challenged Host" or the name of a specific animal. Some examples are:

> "Rheological studies of coagulation change in an animal model of pseudomonas aeruginosa sepsia and pulmonary edema"
>
> "Renal infection in the rat after reflux challenge with pseudomonas aeruginosa"
>
> "Comparative response of various mouse strains to intra corneal challenge with pseudomonas aeruginosa"

It is important for the system to understand all of the researcher's questions when doing each search. In many cases, titles relevant to one question will be found while searching on another question. These relevant titles are not necessarily found in response to the question they address. For example, articles on animal models which were found with the "Exotoxin A" search were not found with the "Animal Model" search.

Current systems are much more accurate when specific names are used: exotoxin, pseudomonas, etc., than when concepts are used: animal model, death, infection, etc.

III. PRODUCT PREPARATION AND EVALUATION

A. Virulence of Exotoxin A

1. Relevance. Of three retrieved articles only one was judged highly relevant by the system. The article reviewed a 10-year history of research on both diphtheria toxin and exotoxin A from P. aeruginosa covering three topics of interest: mechanisms of entry, attachment, and action.

2. Product format/content. The information product had a 4-part format as follows:

   - A statement about the kind of article it was (review) the organisms compared, and the topics of the comparison.
   - The author's abstract.
   - Diagram excerpted from the text comparing diphtheria toxin and endotoxin A from P. aeruginosa in activation (one page).
   - Two-paragraph statement excerpted from the text which suggested differences between the two pathogens in entry mechanism; similarity between the two in method of action, i.e., inhibition of protein synthesis.

3. User evaluations of product. Researcher's evaluation of the product was to praise the first section. He liked the gross characterization of the article, i.e.,

   - A review article
   - Comparing x and y
   - On topics a, b, c.

B. Virulence of Protease, Elastase and Hemolysin

1. Relevance. The system judged three articles to be of very high relevance. Each of the three provided supporting experimental evidence that a high titer of hemolysin was characteristics of virulent strains of P. aeruginosa.

2. Product format and content. Format of the products was a six-part outline:

- Kind of study - e.g., experimental, animal cells
- Rationale - why the author did the study
- Strains of P. aeruginosa and enzyme products studied - a list.
- Methods of assay - names of methods used or details of unlabeled methods.
- Results - written statement of main findings with selected tables and graphs which support each result.
- Conclusions - author's conclusion or its paraphrase.

3. User evaluations of products. Researcher was most interested in the rationale section of each product. He wanted to know why the author did the study, what were the ideas that led the author to do the study?

Researcher also liked the Results format. A statement of a result accompanied by a table or figure on which the statement was based.

Researcher stated that given the products he would not need the full texts.

C. Death from Infection

1. Relevance. The system judged one article 'very high', two articles 'high', one article 'borderline' and two articles of 'low' relevance. The 'very high' judgment was based on the good fit between researcher's stated interests, (i.e., sub-questions) and material presented in the article.

- Described the primary illness of patients.
- Described the sources of bacteria found in the hosts prior to bacteremia.
- Described autopsy report on those patients who died with bacteremia.

The two 'high' relevance judgments were articles of less than complete fit to the user's sub-questions. The 'borderline' article was in French and it was difficult to determine whether the 'host' was an animal or a tissue culture. The 'low' article was statistical and historical and did not tell about 'why they died'. Researcher evaluated the statistical study as 'very high' on the basis of problem importance; article helps justify research on the problem.

The second 'low' relevance article told about hospital procedures to reduce likelihood of infection by secondary pathogens.

2. Product format and content. Format of these products was a page synopsis with 'pointers' to supporting material in the text. Back-up pages to the synopsis contained the 'material-pointed-to' excerpted from the text.

3. User evaluation of products. Products were judged valuable as a screening device, i.e., researcher could use the synopsis and excerpts to decide whether or not he wanted to see the full text. This format did not substitute for the article but enabled the researcher to do a selective follow-up.

D. Animal Models

1. Relevance. 'System' judged three articles of 'very high' relevance since each named an animal model or specified a compromise procedure used with an experimental animal. It would have helped had 'system' gotten from researcher a list of those animal models he knew about so 'system' could recognize a 'new' model.

One article was judged 'moderate' by the system since an animal model was suggested but not specified and article was in German.

2. Product format and content. The format of the products for Animal Models was the same as for Death from Infection. A one-page synopsis of the study was written by a reviewer who selected the points to be included in the synopsis. Mostly a general overview of what was done, how and with what result. The synopses were not excerpts or paraphrasings of the author's writings but the reviewers digest of the article. They were highly individualistic and unstructured products.

Each synopsis was followed by pages of results excerpted from the article. Some with written descriptions and graphics; others with the graphics only.

3. User evaluations of products. The three specific animal models discussed in the 'very high' relevance articles were new to him, he wanted to learn more about each and wanted citations to prior work of the author's wherein the procedure had been described. The 'system' could have made these prescriptions available had it known in advance that the researcher was not familiar with the models named in the articles retrieved. Researcher needed to know details of the procedure for each model in order to decide whether or not it would be feasible to perform in his laboratory.

Dr. James Coolbaugh
Medical Microbiology Branch
Naval Medical Research
Institute
Bethesda, Maryland

I. DIAGNOSIS: CONTEXT AND QUESTIONS

A. Context for Identifying Questions

1. Statement of the research topic. The microbial hazards of polluted waters; specifically the virulence factors in three newly identified genera of enteric microorganisms: aeromonas, campylobacter and yersinia.

2. Statement of the topic's importance; a justification of the research: Navy divers work in polluted waters and may be at risk from the presence of these organisms.

3. Statement of the goal structure:

   - The development of an antidote for or an immunizing vaccine against the above three enteric pathogens.
   - The identification and explication of virulence factors, i.e., how do these organisms cause diarrheal disease.

4. Characterization of the researcher's experience/level of sophistication on the topic: Researcher has been sampling polluted waters, looking for the organisms, and identifying them in these samples. He has not begun to work on their characteristics (enzyme product characteristics) or to experiment with them to determine virulence factors.

5. Characterization of the status of research on the topic

   a. Some facts about the content.

      - The three genera are newly-recognized enteric pathogens.
      - Species and subspecies known to be pathogenic to humans are as follows:
        - Aeromonas
          - hydrophila
          - sobria

- Campylobacter fetus (subspecies)
  - jejuni
  - intestinalis
- Yersinia
  - enterocolitica

- They have been detected in polluted waterways: Norfolk Harbor, New York Harbor, Anacostia River.
- They cause diarrhea, colitus and cholera-like disease symptoms.
- They produce enzymes which are in some way responsible for their virulence/pathogenicity. Their enzyme products include:
  - Hemolysin
  - Protease
  - Elastase
  - Cytotoxin
  - Enterotoxin

b. Some facts about method of study

- Most of the work has been clinical case studies and therefore descriptive. Very little experimental work with the organisms.
- Enzyme products have been tested with sheep and rabbits.
- The 'ileal loop' test using rabbits is a particularly unreliable, inefficient, and procedurally messy test for enzyme effects.

6. Statement of excluded topics

- Campylobacter fetus, subspecies fetus pathogenic to animals only; is not a hazard to divers and, therefore, not of interest to the research.

7. Characterization of researcher's current problem

It has been established clinically that these organisms cause diarrheal disease but there is very little evidence about how they do this, i.e., what are the virulence factors?

B. Questions

The system had recently encountered the 'virulence factors' model with Dr. Pavlovskis and proceeded to structure the questions according to the matrix model of roles-by-enzymes for each genus of microorganism.

1. Virulence factors in aeromonas

a. How is aeromonas transmitted from polluted water to the diver's body?

- Where has aeromonas been found in divers?

b. How does the organism attach itself to host cells?

- Are pili instrumental to attachment?

c. Which enzymes produce the symptoms?

- By what mechanism (mode of action) of the enzyme does disease come about?
- What concentration (infectious dosage level) of the enzyme is required to produce toxic effects?

d. What tests, besides 'ileal loop', have been used to detect and measure toxic effects?

2. Virulence factors in campylobacter

a. How is campylobacter transmitted from polluted water to the diver's body?

- Where has campylobacter been found in divers?

b. How does the organism attach itself to host cells?

c. Does campylobacter produce any enzymes other than those known to be produced by aeromonas?

- Which enzymes cause disease?
- By what mechanism or mode of action?
- What defines an infectious dosage?

d. What tests are used to detect and measure toxic effects?

3. Virulence factors in yersinia

a. How is yersinia transmitted from polluted water to the diver's body?

- Where has yersinia been found in divers?

b. How does the organism attach itself to host cells?

c. Does yersinia produce any enzymes other than those known to be produced by aeromonas?

- Which enzymes cause disease?
- By what mechanism or mode of action?
- What defines an infectious dosage?

d. What tests are used to detect and measure toxic effects?

C. Question-Specific Context (Knowledge Structures Required for Answer-Seeking)

1. About virulence factors in aeromonas

- Aeromonas have been found in polluted waters
- Some strains of aeromonas have pili, others do not.
- Pili have been related to attachment in other organisms.
- Aeromonas produces the following enzymes:
  - hemolysin
  - protease
  - elastase
  - cytotoxin
  - enterotoxin
- A test for enterotoxin is the 'ileal loop' test used with rabbits.
- 'Ileal loop' test is procedurally messy and the results unreliable.
- Various toxins have been tested with human, rabbit and sheep cell cultures.

2. About virulence factors in campylobacter

- Campylobacter is difficult to grow (culture), requires high temperature.
- Campylobacter has not been found in natural water environments, only clinical samples.
- Campylobacter causes diarrhea, and acute colitis, but mechanisms are unknown.
- Enzyme products of campylobacter are not well specified.

3. About virulence factors in yersinia

- Yersinia is difficult to grow (culture)
- Yersinia can cause diarrhea and colitis.

II. SEARCH STRATEGY FORMULATION AND IMPLEMENTATION

Dr. Coolbaugh was interested in finding out about the toxic products (enzymes) produced by five bacterial species, the pathogenic effects of these products, and the tests and procedures used to study them. He also wanted to know how the five bacteria attach to the host cells. The five bacteria were:

- Aeromonas: species hydrophila and sobria
- Campylobacter: subspecies jejuni and intestinalis (both species fetus)
- Yersinia: species enterocolitica

The search strategy was to pair each bacteria with a list of toxic products and a list of words related to attachment. Dr. Coolbaugh mentioned the following toxic products: cytotoxin, entero toxin, hemolytic toxin, protease, elastase. Two data bases were used: BIOSIS and MEDLINE. BIOSIS was selected because it indexes articles on bacteria; MEDLINE was chosen because of the clinical effects of the selected organisms on humans.

A. Aeromonas Hydrophila; Aeromonas Sobria

The first search was on aeromonas hydrophila and sobria and attachment. The search terms are as follows:

BIOSIS	MEDLINE
Aeromonas hydrophila or	Aeromonas hydrophila or
Aeromonas sobria and	Aeromonas sobria and
Attach or	Attach or
Hold or	Adhere or
Pili	Hold or
	Pili

The BIOSIS search led to one article which was written by Dr. Coolbaugh. The MEDLINE search identified one article which was about the association of the ciliate epistylis of large-mouth bass and aeromonas hydrophila. This article was not selected as the title said nothing about attachment or adherence to cells.

The second search was on aeromonas hydrophila and sobria and the toxic products mentioned by Dr. Coolbaugh. No specific searches were done on methods of study as the data bases are not indexed to obtain this information. The search terms were as follows:

BIOSIS	MEDLINE
Aeromonas hydrophila or	Aeromonas hydrophila or
Aeromonas sobria and	Aeromonas sobria and
Entero (w) toxin or	Hemolysins or
Hemo (w) lysin or	Cytotoxins or
Protease or	Pancreatopeptidase
Elastase or	
Cyto (w) toxin	

The BIOSIS search resulted in 23 citations; MEDLINE identified seven. Three articles were identified by both data bases. There were 27 new citations altogether. The system selected 20 of these as being relevant. Citations were rejected for the following reasons:

- They were on general methods: "Methods of enzymology, Vol. XLV. Proteolytic enzymes"
- They were about antigens: "A study of the antigenic structure of non-agglutinating vibrios and bacteria of the genus aeromonas"

Selected citations either mentioned "extra cellular enzymes", a specific enzyme, or pathogenic products. Some examples are:

- "Characterization of 3 aeromonas-spp and 9 pseudomonas-spp by extracellular enzymes and hemolysins"

- "Entero toxins of enteric bacteria: A review" (Note: Aeromonas are enteric bacteria)
- "Cyto toxic entero toxin produced by aeromonas hydrophila: Relationship of toxigenic isolates to diarrheal disease"

B. Campylobacter Jejuni, Intestinalis (Species Fetus)

Very little has been published on Campylobacter Jejuni or Intestinalis: BIOSIS showed 4 articles. These were all printed out. Although none was exactly on the topic, we selected all 4 as potentially useful. The system judged that the articles might contain something about the products of these bacteria. Next in BIOSIS the system tried Campylobacter Fetus - this gave 113 citations.

- Campylobacter fetus and attach or hold or adhere or pili resulted in 0.
- Camplyobacter fetus and extoxin or hemolysin, etc., resulted in 2 citations, both of which looked like general reviews (e.g.): "Mechanism of Bacterial Diarrheas"

They both were selected for further review.

MEDLINE identified 27 citations on campylobacter fetus and one specifically on toxic products. Twenty citations were selected. These citations had one of the following characteristics:

- Included jejuni or intestinalis
- Indicated campylobacter as a factor in some human disease.

Some examples are:

- "Acute colitis caused by campylobacter fetus ss jejuni"
- "Campylobacter entretis: A common cause of adult diarrohoea"
- "Acute colitis and bacteremia due to campylobacter fetus"

C. Yersinia Enterocolitica

Searches were the same as for aeromonas. In BIOSIS, no articles were found on attachment; in MEDLINE, two were identified. One was about serum antibodies, the other was on plasmid-mediated

tissue invasivesness. Neither citation mentioned anything about attachment; both were rejected.

The searches on toxic products resulted in 12 citations from BIOSIS and 9 from MEDLINE; 5 of these citations were the same giving a total of 17 new citations. Ten citations were selected. Selected citations had the following characteristics:

- Included a toxin name
- Related yersinia to other bacteria
- Included the word "pathogenesis"

General Comments

Dr. Coolbaugh's questions were similar to those raised by Dr. Pavlovskis about virulence factors and their effects. The search results, however, show a large difference in the level of work that has been accomplished. The work in pseudomonas aeruginosa is much more detailed and is further down the line in examining characteristics of the toxic products (e.g.: how they work, what they are made up of, etc.) The articles on aermonas and toxic products are all at the level of determining which products are produced. The citations about campylobacter and yersinia showed even less progress. Much of the work with campylobacter described the organism; work with yersinia was to link the organism with a specific disease.

III. PRODUCT PREPARATION AND EVALUATION

A. Aeromonas

1. Relevance

The system prepared five products from articles about aeromonas hydrophila. Each article addressed either the issue of attachment or the characteristics of the enzyme products of aeromonas hydrophila. The system made no distinctions among the articles on degree of relevance.

2. Product format and content

Each of the five products was prepared as compressed substitutes for the original document. Headings were used which followed standard journal format, i.e.,

- Problem
- Strains of organism studied
- Methods
- Results

The problem statement paraphrased the article, strains were listed, methods were mostly excerpts and results were statements of main findings supported by tables/graphs excerpted from the article.

3. User evaluation

The researcher found these products too detailed for his present state of progress on the research topic. He was just beginning to shift from in-the-field sampling for aeromonas to planning laboratory experiments with the organisms.

The system had noted this difference in state of progress between Coolbaugh and Pavlovskis but failed to account for its effect on product requirements. The system prepared products

as though they were to be used by Pavlovskis, used a format/content strategy which had been successful with Pavlovskis (i.e., a miniature of the original document).

Coolbaugh found the products telling him things he was not ready to absorb/make use of; maybe later they will have an impact on his work but not now. Very detailed excerpts from the articles did not connect with any specific problem he has as yet. Now, at the initial planning stage of experimental research with aeromonas, Coolbaugh wanted an overview of what was being done with aeromonas, i.e., the status of experimental work and not how it was being conducted. He suggested to the system formats/contents more appropriate to his needs at this stage of progress.

- Give the author's rationale in a short statement. Why did this researcher do this work?
- List the topics treated in the paper.
- List the results; do not reproduce the results in excerpted tables, etc., but state, like a list, what the tables tell.
- Limit the product to a single page.

The resulting products should serve two functions for the researcher:

a. Enable him to acquire an overview (at the level of what) of the status of experimental research with aeromonas; who's doing what, experimentally with aeromonas.

b. Provide him with a means to selectively access details of the studies at a later time.

The researcher suggested that the idea of the product is to go beyond the 3 x 5 citation card so as to have available an overview of the article in the form of lists.

B. Aeromonas and Campylobacter

The system prepared products from five articles: two about aeromonas and three about campylobacter.

1. Relevance

System judged the five articles using a three-category scale: high/medium/low. The high rated articles dealt with attachment or enzyme products or methods of testing for toxicity of a given enzyme product. The lower rated articles were about the organism but dealt with its morphology or taxonomy. The researcher was in general agreement with the system's ratings except for those campylobacter articles rated low by the system. The researcher knows so little about campylobacter that even articles about taxonomy/morphology were judged highly relevant.

2. Product format and content

All articles were paraphrased not excerpted. All were of a page in length. All were unstructured in format, i.e., no headings, but contained structured information.

Products contained a statement of purpose/rationale, a list of organisms studied, lists of methods used, lists of topics addressed in the results.

Products contained no excerpts, no tables, graphs or figures.

3. User evaluation

The researcher, having specified the format, content and style which he preferred, was most pleased with these products.

Perhaps our system should provide the user with a 'menu' of formats, etc., to choose from, as a way to help the system provide acceptable products.

Dr. Emilio Weiss
Medical Microbiology Branch
Naval Medical Research
Institute
Bethesda, Maryland

I. DIAGNOSIS: CONTEXT AND QUESTIONS

A. Context for Identifying Questions

1. Statement of the research topic: the problem of developing a non-cell growth medium for rickettsia.

2. Statement of the topic's importance; a justification of the research: rickettsia is a genus of microorganisms which causes Rocky Mountain spotted fever and epidemic typhus.

3. Statement of the goal structure:

••• To develop an immunizing vaccine against rickettsia.

•• To develop a non-cell growth medium so that rickettsia can be easily grown in large numbers for experimentation.

• To identify compounds which rickettsia can use for energy (can metabolize).

4. Characterization of the researcher's experience/level of sophistication with the topic: Dr. Weiss is an expert in research on the family of microorganisms which includes the genus rickettsia.

5. Characterization of the status on the topic:

a. Some facts about the content -

- Rickettsia can only be grown in a host-cell medium; they are difficult to grow and slow to multiply. This slows the rate of experimentation with the organism.
- The family rickettsiaceae includes three genera: rickettsia, coxiella and rochalimaea.
- Rochalimaea has a DNA structure similar to rickettsia.
- Rochalimaea can be grown in a cell-free medium.

b. Some facts about methods/approaches:

- Prior work toward identifying compounds which rickettsia can metabolize has been of a shotgun character, and has been unsuccessful.

6. Statement of excluded topics

- The researcher is not interested in legionella.
- The researcher is not interested in enteric bacteria or in gram-positive cocci.

7. Characterization of the researcher's current problem

- Use rochalimaea, which are plentiful, as a simulation of rickettsia. Try a selected set of four biochemical reactions with rochalimaea. If rochalimaea tests positive to any of these reactions, then try them (it) with rickettsia.
- Dr. Weiss has selected three strains of rochalimaea as the experimental organisms, and four biochemical reactions as tests for positive indication of metabolism by the experimental organisms:
  - Ornithine Metabolism
  - Voges-Proskauer Reaction
  - Malonate Inhibition of Succinate Metabolism
  - Hematin Metabolism

Dr. Weiss' expectation is that the experimental organisms will test positive to one or more of these biochemical reactions.

B. Questions

Before implementing his planned series of tests, the researcher decided to check the literature for recent evidence of positive reactions to these four 'tests', for any improvements in technique for studying each 'test' and for information about current indices/measures used to detect and quantify the biochemical reaction. The system organized these information requirements into four question sets as follows:

1. Question Set #1. About Ornithine Metabolism

   a. What is the evidence of positive reaction for ornithine metabolism by any gram-negative, non-enteric, microorganism?

   b. What are the procedures for studying ornithine metabolism?

   c. What are the indices/measures used to detect and quantify ornithine metabolism reaction?

2. Question Set #2. About Voges-Proskauer

   a. What is the evidence of positive reaction for Voges-Proskauer by any gram-negative, non-enteric, microorganism?

   b. What are the procedures for studying Voges-Proskauer?

   c. What are the indices/measures used to detect and quantify Voges-Proskauer reaction?

3. Question Set #3. About Malonate Inhibition

   a. What is the evidence of positive reaction to malonate inhibition of succinate metabolism by any gram-negative, non-enteric microorganism?

   b. What are the procedures for studying malonate inhibition of succinate metabolism?

   c. What are the indices/measures used to detect and quantify malonate inhibition of succinate metabolism?

4. Question Set #4. About Hematin Metabolism

   - Is there any evidence that any microorganism metabolizes soluable ferric pyrophosphate?

C. Question-Specific Context (Knowledge Structures Required for Answer Seeking)

1. About Ornithine Metabolism

   a. Ornithine is an amine; an intermediate product in the breakdown pathway between amino acid and urea.

   b. Other compounds in this breakdown pathway include alanine, proline, carbamoyl phosphate, citrulline and arginine.

   c. The family of microorganisms of interest (Rickettsieae) are gram-negative, non-enteric microorganisms.

   d. Evidence of positive ornithine metabolism by any gram-negative non-enteric microorganism would be of interest.

2. About Voges-Proskauer

   a. The primary breakdown pathway is the conversion of glucose to pyruvate to acetoin.

   b. The index of a positive Voges-Proskauer reaction is the presence of acetoin.

   c. Acetoin is also called aceylmethylcarbonyl; also 3-hydroxy, 2-butanone

3. About Malonate Inhibition

   a. Malonate is a compound known to inhibit the ability of an organism to metabolize succinate.

   b. Succinate may be an important metabolite for rickettsia.

4. About Hematin Metabolism

   a. Hematin is an organic iron compound found in blood hemoglobin.

   b. Most microorganisms metabolize hematin and rickettsia requires an iron compound for growth.

   c. Ferric pyrophosphate is a non-organic iron compound.

d. Ferric pyrophosphate was detected in blood samples from victims of Legionnaire's disease.

e. Legionnaire's disease may have been caused by a microorganism of the rickettsiaceae family.

f. Rickettsia may be able to metabolize pyrophosphate as a substitute for hematin.

II. SEARCH STRATEGY FORMULATION AND IMPLEMENTATION

Dr. Weiss is studying rochalimaea as a model for ricketsia. He is interested in determining what compounds rickettsia metabolizes. He has four compounds in mind:

- Ornithine
- Acetoin (other words are pyruvate, Voges Proskauer, acetyl-methyl-carbinol, and 3-hydroxy-2-butanone)
- Malonate inhibition of succinate
- Hematin, ferric pyrophosphate.

He had three questions about each of these compounds:

- What gram negative, non-enteric bacteria have had positive reactions to the compound and what was the reaction?
- What new methods/techniques have been used to study reactions to the compound?
- What are the current indicies used to measure each reaction?

The search strategy was to combine non-enteric, gram-negative bacteria with metabolic processes and with each compound. Both BIOSIS and MEDLINE were searched. BIOSIS indexes citations on bacteria names; MEDLINE indexes clinical literature on bacteria names. In the current search, MEDLINE identified more relevant articles than BIOSIS. Terms on methods of measurements were not used as they are not very useful in identifying articles on methodology. The idea was that method would be identified through a review of the articles.

A. Ornithine Metabolism

Search terms were selected for classes of bacteria which were gram-negative and non-enteric. There may have been some problems with this as the system did not know all of the types of bacteria which are both gram-negative and non-enteric. The present structure of the data base does not allow/provide for sorts of gram-negative or non-enteric.

The second set of search terms were those codes which deal with processes like metabolism. There are other processes contained in these same codes. It is not possible to get just articles on metabolism. This creates a problem when reviewing titles for relevance. If the concept of metabolism is not in the title there is no way of knowing if it is in the article as opposed to some other process under that code. In these cases the system must review abstracts and descriptors.

The third descriptor was the compound name: ornithine. All three descriptor sets were combined with ands for both BIOSIS and MEDLINE. In BIOSIS, two articles were identified, in MEDLINE six articles were found. Six of the eight articles were selected by the system as being relevant. Selection decision was based on whether the title included the word ornithine or any compound in the ornithine pathway (e.g.: arginine, citrulline, etc.). The rejected titles did not contain any of these compound names. Some examples of selected titles were:

- "Poising of the arginine pool and control of bio-luminescence in Beneckea-harveyi"
- "Arginine biosynthesis in Neisseria gonorrhoeae: enzymes catalyzing the formation of ornithine and citrulline"

It was not clear to the system whether these articles were on target (e.g.: were about metabolism, were about non-enteric, gram-negative bacteria.

B. Malonate/Succinate

In order to retrieve articles on the inhibition of succinate metabolism by malonate the system asked for all articles which included:

- Malonate
- Succinate
- Codes for non-enteric, gram-negative bacteria

This combination led to no articles. The system broadened the search by leaving out succinate but no articles were identified. Finally succinate with bacteria was tried: four citations resulted. Two citations were selected as relevant because they included the word succinate and the titles implied inhibition. One of the citations included the word malonate. An example of a rejected title is:

> "Treatment of otitis media caused by halmophilus-influenzae: Evaluation of three anti-microbial regimens"

The selected titles were:

- "Synthesis of alpha keto glutarate by reductive carbonylation of succinate in veillanella selenomonas and bacteroides species"
- "Reversal of succinate-mediated catabolite repression of alkyloulfatase in pseudomonas aeruginosa by 2, 4-dinitrophenal by sodium malonate"

C. Pyruvate/Acetoin

The search terms included the following combinations:

- Voges Proskauer (this is the name of a reaction that changes pyruvate to acetoin) and non-enteric, gram-negative bacteria.
- Pyruvate or acetoin and gram-negative, non-enteric bacteria.

In BIOSIS, Voges Proskauer and bacteria led to no citations. In MEDLINE 10 were identified. Pyruvate or acetoin and bacteria resulted in 18 citations from BIOSIS and 12 from MEDLINE. Fourteen citations were selected as relevant. All those listed under Voges Proskauer were selected. Others were chosen because they contained one of the following words:

- Pyruvate
- Acetoin
- Acetyl-methyl-carbinol
- 3-hydroxy-2-butanone

Still others were chosen because they implied cellular uptake rather than metabolic products (the latter were rejected). Some examples of selected titles are:

> "The pathway of formation of acetate and succinate from pyruvate by bacteroides succinogenes"
>
> "Preliminary crystallographic study of omega-amino acid: pyruvate amino-transfer are from pseudomonas sp. P 126"
>
> "Rapid test for acetyl-methyl-carbinol formation entero-bacteriaceae"

D. Hematin, Ferric Pyrophosphate

There were no articles on ferric pyrophosphate and bacteria. In this search all bacteria were included. Next, iron or pyrophosphate and bacteria was tried. In BIOSIS, one citation was identified:

> "A chemically defined medium for growth of legionella-pneumophila"

This citation was not selected as Dr. Weiss said he knew all about legionella. In MEDLINE, 19 citations were identified. The MEDLINE search also included the term enterochelin. This term was given to the system by Dr. Weiss. Sixteen of the 19 citations were selected as possibly relevant. All of these citations contained the word iron or some synonym (e.g.: Fe, ferric, enterochelin, cytochrome, sider). Some selected titles are:

- "Enzymatic hydrolysis of enterochelin and its iron complex in escherichia coli K-12"
- "Feasibility of enterochelin as an iron-chelating drug: studies with human serum and a mouse model system"
- "Enterochelin (enterobactin): virulence factor for salmonella typhimurium"

The system was unable to judge from the titles whether the articles were about metabolism.

General Comments

The system did not have an adequate frame for interpreting the computer output and determining its relevance to the questions.

The system needed more content context than a few key words for each question.

- All the synonyms for the compounds, processes and bacteria under question.
- What the question really meant - e.g., what is metabolism, what is pyruvate, what are the contexts in which pyruvate is used which are not of interest, which are of interest, etc.?

Another problem encountered further down in the analysis process is that the system had no criteria for determining whether a method was a new method or an old method (one he already knew about). We did not know enough about what he knew to enable us to recognize new knowledge.

The system should have gotten more question-specific context so it could distinguish known from unknown, and so it could recognize relevant citations.

- Name some "gram-negative, non-enteric organisms".
- Specify the tests for/evidence of 'positive reaction'.

III. PRODUCT PREPARATION AND EVALUATION

A. Ornithine Metabolism

1. Relevance. Six articles were retrieved and reviewed. All appeared relevant in that some microorganism was identified as having some effect on a breakdown pathway that included ornithine. For example, the first article tells about how a microorganism (agrobacterium radiobacter) of a certain characteristic (having a nopaline ti-plasmid) degrades arginine via ornithine to generate a carbon which enables bacterial growth. Five of the six articles included complex figures which appeared to describe the breakdown sequence in chemical structure notation. During the review, the system learned the chemical structure of ornithine, arginine and other words that the researcher had talked of. For example, ornithine is also

$$H_2N - (CH_2)_3 - \underset{\displaystyle NH_2}{\underset{|}{CH}} - COOH$$

The system could not discriminate levels of relevance among the six articles.

2. Product format/content. All products had synopsis-plus-backup-detail format. The synopses were very short; a few sentences which gave the sense of the article. For example, "this article presents evidence that x organism contains y enzyme which converts A to B." The backup material consisted entirely of excerpts from the original article. Any pathway diagram was included; also results tables/figures, in one case the entire Methods section where this was the main contribution of the article, i.e., methods of assay for detection and quantification of ornithine metabolism.

3. User evaluation of products. The researcher first reviewed the list of six citations selected by the system as relevant to ornithine metabolism. He judged five of the six as relevant; one was judged to be irrelevant as the title suggested an emphasis on physical chemistry rather than biochemistry.

Next the researcher reviewed the synopsis and excerpted material and decided that the five were clearly relevant and of interest, and confirmed that the sixth was physical chemistry and irrelevant. He was most pleased to see the biochemical breakdown pathways excerpted in the products.

The researcher was then given the full documents which he briefly scanned. His evaluation of the overall procedure was that the synopses and excerpts were of marginal benefit as long as the full text was available. He judged that he could scan the document as easily as he could the excerpts.

Remaining information products were reviewed in the same way and with the same outcome.

Bruce Gunn
Marine Microbiology Lab
University of Maryland

I. DIAGNOSIS: CONTEXT AND QUESTIONS

A. Context for Identifying Questions

1. Statement of the research topic:

The speciation of streptococcus and staphylococcus isolates from aquatic environments; comparisons with clinical species.

2. Statement of the topic's importance; a justification of the research:

Strep and staph in aquatic environments may pose a threat to the health of humans and to commercially valuable marine organisms (oysters, clams, crabs, fish, etc.)

Strains/species of strep and staph may be useful as indicators of fecal pollution in aquatic environments.

3. Characterization of researcher's experience/level of sophisication on the topic: The researcher is an expert in taxonomy of clinical species of strep. He has conducted all phases of work in the taxonomy research paradigm for clinical isolates. However, he is new to the field of aquatic microbiology.

4. Statement of excluded topics:

- Researcher is not interested in clinical strains of staphylococci and streptococci.
- Researcher is not interested in fresh water strains.
- Researcher is not interested in streptocci from animal feces.

5. Characterication of the researcher's current problem:

He has essentially finished that half of the project which will yield clusters of clinical strains of strep for comparison with aquatic strains.

He is just beginning the aquatic half of the problem, i.e., Problem Selection in the research paradigm.

B. Questions

1. Question Set #1: About the current status of taxonomy studies with aquatic strep or staph.

Asked
- What species or strains have been identified in natural bodies of water, particularly Chesapeake Bay?

2. Question Set #2: About the use of strep or staph as indicators of fecal pollution.

Asked
- Is there evidence of the use of strep or staph as an indicator of fecal pollution?

Asked
- Can human vs animal sources of fecal pollution be discriminated in terms of species of strains of strep/staph?

3. Question Set #3: About isolation methods used with aquatic samples.

Asked
- What is the evidence for the value of the membrane filter test for fecal strep?

Asked
- Are there other methods in use for isolating strep/staph from aquatic samples, and how do they compare with Membrane Filter Test in effectiveness re: isolating strep or staph?

Unasked (Name) ∘∘ What isolation media are used?

Unasked (Description ∘∘∘ What is the composition of the medium? of compounds)

Unasked (Preparation ∘∘∘∘ What is the procedure for preparing the procedure) medium?

C. Question-Specific Context for Recognizing Answers

1. Question Set #1: About taxonomy of aquatic strep or staph

SL
- Taxonomy is sorting organisms into hierarchical sets.
  - • Family
  - •• Genus
  - ••• Species/strain

R
- Strep and staph are genus names

R • Streptococcus is also known by the following:

- Gram-positive cocci
- Enterococci

• Subgroups of streptococci of interest to the researcher are:

- Group D
- Viridans

SL - Facales

R • Staphylococcus is also known as micrococci

R • Staphylococcus is a genus of the family Micrococcacae

R • Aquatic in this case means saline water as in Chesapeake Bay; not fresh water as in lakes.

2. Question Set #2. About strep/staph as indices of fecal pollution

R • Strep/staph are gram-positive cocci.

R • The typical index of fecal pollution is some gram-negative rod such as E. coli.

3. Question Set #3. About isolation methods.

R • The Membrane Filter Test is a method for isolating strep from a sample of water polluted by feces.

R • The Membrane Filter Test is not reliable, tends to yield varying results.

SL • Other criteria for assessing isolation methods (in addition to reliability) include:

- Simplicity/ease of application
- Availability of materials and equipment requirements.

SL • There are other methods for isolating strep or staph (researcher knew two others, but system failed to ask for specification).

SL • All methods of isolation include a medium of compounds designed to select for the organism to be isolated. These are called 'recovery media' or 'selective media'.

SL • A selective medium is composed of compounds which the organism uses for energy (metabolizes).

SL • A medium is sometimes called 'agar', or 'broth', i.e., PSE media is the same as PSE agar and PSE broth.

- Effectiveness of a medium is the extent to which it captures the organism of interest and no others.

SL

- Different recovery media (agars) can be used with the Membrane Filter Test.

SL

- Different recovery media (broths) can be used with the most probable number test (MPN).

II. SEARCH STRATEGY FORMULATION AND IMPLEMENTATION

The researcher was working on the taxonomy of streptococci from brackish and sea water. He had three major question areas:

- What taxonomy has been done on streptococci and staphylococci in an aquatic environment?
- What is the significance of gram-positive cocci in the aquatic environment as indicators of fecal pollution? Do human feces contain different species than animal feces?
- What isolation methods for streptococci have been used in aquatic environments other than the membrane filter test? Who started the membrane filter test and how effective is it?

The search strategy was to pair up all the names for streptococci with the key words in each question set. The BIOSIS data base was used as it indexes most of the literature on bacteria, particularly environmental strains of bacteria.

A. Taxonomy

The search on taxonomy of aquatic streptococci contained the following terms:

- Streptococci
- Staphylococci
- Enterococci
- Micrococci
- Peptococci

— the organism

- Oceanography
- Limnology

— aquatic environment

- Taxonomy - bacteriology general
- Taxonomy nomenclature

Overall 51 citations were identified; 15 of these were selected as relevant. Most selected titles (13 of 15) included some word which implied classification: taxonomy, speciation, characteristics,

differentiation of species, identification by some test. All selected titles included streptococci. One title, not using classification terms, identifies a specific characteristic of streptococcus:

"Pyruvate fermentation by streptococcus facalis"

The system recognized "pyruvate fermentation" as a characteristic and then inferred that if a characteristic were mentioned then some classification/taxonomy would be involved. Taxonomy is based on characteristics. "Fermentation" was recognized as a process that the organism does – thus, a characteristic of the organism.

The other title (without classification words) was "Group D Streptococci". The system reasoned that a title which only stated the name of a group of organisms would be about the characteristics of that group of organisms. Since group D was mentioned by the researcher, this article was selected.

Examples of other selected titles are:

"Indicator organisms - A review. I. Taxonomy of the fecal streptococci"

"Characterization of two new isolates of beta hemolytic streptococci from Amazon fresh water dolphins inia-geoffrensis"

"Species differentiation of group D streptococci"

"Presumptive speciation of streptococcus bovis and other group D streptococci from human sources by using arginine and pyruvate tests"

The titles that were not selected either did not imply classification or did not specifically mention streptococci. Additionally, in many cases, the source of the bacteria was from animal waste. The researcher was mainly interested in streptococci associated with humans. Some examples of rejected titles are:

"Effects of upgrading a municipal waste water effluence on pollution indicator and other microorganisms in river water"

"A study of bacterial flora isolated from marine algae"

"The incidence of enteric bacteria in the sediment of a bird sanctuary pond"

"Microbial investigation of the high mountain vevcani lake"

P. Significance of Streptococci in the Water

The search terms for the significance of fecal streptococci in aquatic environments were:

- Streptococci
- Staphylococci
- Enterococci
- Micrococci
- Peptococci

} – the organism

- Oceanography
- Limnology

} – aquatic environment

- Fecal
- Feces

Twenty of the most recent articles were printed out. Of these nine were selected as having relevance to the question. Selected titles contained the following words.

- Indicators of contamination, pollution
- Distribution of fecal bacteria
- Persistence, survival of fecal bacteria
- Sanitary significance of streptococci
- Bacterial, microbial water quality

Titles implied the possible presence of streptococci from human sources in natural waters (e.g.: sea water, rivers and streams, specific bodies of water). Some examples of selected titles are:

"Fecal bacterial contamination of trout hatchery water effluent"

"Survival of fecal streptococci in sea water"

"Pollution indicators and other microorganisms in river sediment"

"Types and sanitary significance of fecal streptococci isolated from feces, sewage and water"

Twenty additional articles on the significance of streptococci in the water were identified through the review of bibliographies in the selected articles.

Rejected citations were on analyses of water supplies or on the impact of birds and animals on the level of bacteria in the water.

C. Methods of Isolation: The Membrane Filter Test

This search included all the terms for streptococci and water as well as microbial ultrastructure, membrane filter, isolation. Ten articles were identified and one was selected as useful/relevant. This article was:

"Membrane filter technique for enumeration of enterococci in marine waters"

Rejected articles were mainly about isolating other organisms (pseudomonas, salmonella), or about isolating bacteria from fish or from aquarium plants. Examples are:

"Isolation of salmonellae and other potential organisms from the fresh water aquarium snail ampullaria"

"In vitro studies on drug resistance with microorganisms isolated from marine plants"

Nineteen additional articles were selected by starting with the bibliography of the relevant article and working backwards. All selected articles had one or more of the following terms:

- Media
- Test
- Detection
- Isolation
- Name of a particular method.

Names of additional methods, beyond membrane filter, were identified through reading selected sarticles. These method names were used to do further searching.

III. PRODUCT PREPARATION AND EVALUATION

Products were prepared for Question Set #3, Isolation Methods. There were two separate packages: one about the Membrane Filter Test, and a second about other methods.

A. Membrane Filter Test

1. Relevance. Documents were obtained for ten citations and all were judged relevant by the system. All had to do with the use of various recovery media with the Membrane Filter Test; all included streptococcus as a target organism; all included natural water samples.

2. Format and content. Each article was reduced to a short synopsis of 4-8 sentences prepared by a reviewer. The synopses were not excerpts from the article. The synopsis characterized the article, i.e., told what it was about, and included 'pointers' to amplifying information. For example: "Composition of a new medium, ME agar, is described (p. 591)." The paragraph describing media composition was excerpted from the original and included as a back-up sheet to the synopsis. A typical product included the synopsis and back-up sheets containing one or two paragraphs and one or two tables 'pointed to' in the synopsis.

3. Procedure and user Evaluation

   a. The researcher was shown a list of the 10 citations which the system had selected as relevant. The researcher scanned the list and made three responses.

      - He was overjoyed to see 10 rather than 100 citations more typical of computer search results.
      - He judged all 10 to be highly relevant.
      - He selected citation #8 as the article he would read/review first since the title suggested that eight media procedures were compared for effectiveness; also the date of this citation was the most recent of the 10.

b. Next the researcher read each synopsis and accompanying back-up material as prepared by the system and made the following judgments:

- Citation #9 provided no new information. If he had the other documents, he could ignore this one.
- The remaining citations were valuable to his work and he wanted full documents in order to follow-up details not provided by the synopsis-with-pointers-to-excerpts.

Some of the detailed information he desired but were not in the prepared products were as follows:

- What did they do to avoid clogging the filter with debris?
- What procedural problems are there in using the method in the field vs in the laboratory?
- What compound in the medium selected for strep?
- Why did the technique/medium produce false positives, i.e., recover organisms other than those of interest?

The researcher found the synopses and attached excerpts valuable for determining which articles he should follow-up. The system provided answers to questions it knew about, but the researcher had more detailed questions which he did not tell the system about in the earlier interaction. As the researcher reviewed materials prepared by the system these questions surfaced. In a real-time interactive mode, the system could have 'comprehended' those questions and elaborated upon the products.

Also, not all the questions the researcher needs to ask can be asked by him in an initial need-diagnosing interaction. For example, in this case the researcher came across a recovery medium unknown to him, DSE, named in one of the synopses. Now the researcher had

detailed questions about DSE which had not been previously expressed. The system perhaps could have had the researcher specify all the media known to him in the beginning, then instructed itself to answer detailed questions about any medium it encountered which was not on this list.

B. Other Isolation Methods

1. Relevance. The system had identified citations which discussed five isolation methods in addition to Membrane Filter Test. All were used to isolate streptococuccus from water samples.

2. Format and content. The first product answered the 'what other methods' question at the level of name only.

The second product was a list of citations, by publication year, for each of the methods.

The third product consisted of synopsis with-pointers-to-excerpts for each article within each method-type subset. These products were modeled after those prepared for articles about the Membrane Filter Test. A six-sentence synopsis with pointers to paragraphs, tables, figures, etc., which were excerpted from the full-text document and appended to the synopsis.

3. Procedure and user evaluation

a. First the research reviewed Product 1, a list of 5 names as follows:

A. Most Probable Number (MPN)

B. Pour Plate

C. Precipition Test

D. Coagglutination Method

E. Fluore scent-antibody technique (FA)

The researcher used this product to direct the order of priority of further searching on the basis of those methods he knew least about; in this case the reverse of the order of presentation. The researcher knew least about methods E and D, some of C, more of B and most about A.

b. Next the researcher reviewed the citations listed for Method E and the synopsis with pointers for the articles cited. There were three synopses and researcher decided that two would not be useful as there were no applications of the method, only descriptions. The third he would follow-up since the synopsis suggested that the method had been tried out with some success and that the materials required were commercially available.

c. Next he reviewed the citation list and synopsis-with-pointers for Method D. The one article was of no further interest since it was not a standardized or proven test for use in the field, only in the laboratory.

d. Next he reviewed the citation list and synopsis set for the articles about Method C and came to the same conclusion: the method was not adapted to field use.

e. Now the researcher went back to methods he knew well to look for new developments.

Those articles which suggested lack of new developments he rejected for future follow-up; those which included aspects of interest to his application he noted for future follow-up. 'Good' articles were those which were applicable to aquatic sampling in some way. For example:

- Included water salinity ranges as a factor in recovery. Researcher expressed concern about whether or not MPN could be used in saline waters like Chesapeake Bay.
- Discussed problems in use of method in the field.
- An early paper in MPN of historical significance for use in an INTRODUCTION.